

HELSINGIN YLIOPISTO

A GENRE ANALYSIS OF ACADEMIC RESEARCH BLOGS

Katriina Tiainen
MA Thesis
English Philology
Department of Modern Languages
University of Helsinki
June 2012

CONTENTS

1 INTRODUCTION	1
2 THEORETICAL BACKGROUND.....	6
2.1 GENRE THEORY.....	6
2.1.1 <i>The concept of genre</i>	6
2.1.2 <i>Main goals of genre theory</i>	9
2.1.3 <i>Genre as social action</i>	10
2.2 ACADEMIC GENRES.....	12
2.2.1 <i>Previous research on academic discourse</i>	12
2.2.2 <i>Academic genres and their organisation</i>	15
2.2.3 <i>Linguistic features of academic discourse</i>	17
2.3 COMPUTER-MEDIATED COMMUNICATION AND WEB GENRES.....	18
2.3.1 <i>Genre research on computer-mediated communication</i>	19
2.3.2 <i>Blogs and blog research</i>	21
2.3.3 <i>Academic blogs</i>	23
3 MATERIAL AND METHODS	27
3.1 ANALYSING UNFAMILIAR GENRES.....	27
3.2 DATA COLLECTION	31
3.3 DATA ANALYSIS	31
4 RESULTS AND ANALYSIS	35
4.1 DISCOURSE COMMUNITIES.....	35
4.1.1 <i>Blog author characteristics</i>	35
4.1.2 <i>Audience and reader characteristics</i>	38
4.1.3 <i>Interaction between the author and the audience</i>	40
4.2 COMMUNICATIVE PURPOSES	41
4.2.1 <i>Invisible college</i>	41
4.2.2 <i>Diary about academic life</i>	43
4.2.3 <i>Popularising science</i>	45
4.2.4 <i>Soap box</i>	47
4.3 STRUCTURAL FEATURES.....	49
4.4 LINGUISTIC FEATURES	55
5 DISCUSSION	59
5.1 DISCOURSE COMMUNITY	59
5.2 COMMUNICATIVE PURPOSE.....	61
5.3 STRUCTURAL FEATURES.....	65
5.4 LINGUISTIC FEATURES	67
5.5 ACADEMIC RESEARCH BLOG AS AN ACADEMIC GENRE	68
5.7 RELIABILITY AND VALIDITY	69
6 CONCLUSION	72
REFERENCES	75

APPENDIX: Lexical items for linguistic features analysed in the study

LIST OF TABLES AND FIGURES

TABLE 1: USES OF BLOGS IN ACADEMIC INSTITUTIONS (STUART 2006: 392).....	24
TABLE 2: THE INTERPLAY BETWEEN FUNCTION OF AND MOTIVATION FOR BLOGS USE IN A SCHOLARLY CONTEXT (KJELLBERG 2010)	26
TABLE 3: EXAMPLE OF THE ANALYSIS OF COMMUNICATIVE PURPOSES	32
TABLE 4: GENERAL BLOG AUTHOR CHARACTERISTICS	35
TABLE 5: BLOG AUTHOR TITLE	36
TABLE 6: FIELD OF SCIENCE	37
TABLE 7: INTENDED AUDIENCE.....	38
TABLE 8: READERS.....	39
TABLE 9: INTERACTION BETWEEN BLOG AUTHOR AND AUDIENCE	40
TABLE 10: PRIMARY PURPOSE	41
TABLE 11: STRUCTURAL FEATURES	50
TABLE 12: LINKS	53
TABLE 13: COMPARISON OF LINGUISTIC FEATURES IN ACADEMIC BLOGS WITH BIBER’S STUDY OF ACADEMIC PROSE (1988: 255)	57
FIGURE 1: ACADEMIC COG	2
FIGURE 2: THE CONFERENCE PROCEEDINGS PAPER GENRE CHAIN (RÄISÄNEN 1999: 112).....	16
FIGURE 3: THE EVOLUTION OF CYBERGENRES (SHEPHERD & WATTERS 1998: 2)	20
FIGURE 4: BADGES ON PRACTICAL ETHICS	52
FIGURE 5: BADGES ON LEARNING CURVES.....	52

1 INTRODUCTION

Advances in communication technology have made it increasingly easy for anyone with an Internet access to publish their thoughts online. One of the most popular ways of doing this is to create and maintain a blog, that is, a frequently modified Web page where dated entries are listed in reverse chronological order. The amount of public blogs has been steadily growing since the turn of the millennium, and as of the beginning of 2011, there were more than 156 million blogs in existence.

At the same time, scholarly publishing is undergoing a drastic transformation. Printed scientific journals are becoming increasingly expensive, which has led many libraries to prefer the online versions of the journals. However, because of the growing amount of electronic journals and the ever-rising subscription fees, sometimes even the electronic versions of scientific journals have become too costly for libraries. With the *publish or perish* attitude of the scientific world, researchers are aware that it is not enough to publish an article in a journal that no one can afford to read. This has resulted in Open Access publishing, a practice which provides online access to scientific papers without having to pay for subscription fees. Therefore, much of the scientific publishing takes place on the Web, and has, in the case of Open Access, opened up to anyone with an Internet access. With this in mind, researchers have realised that an active online presence is very important in making a successful academic career.

This understanding of the importance of creating a presence of oneself on the Internet has paved the way for many academics to start writing a scientific blog of their own. In these blogs, academics can disseminate and discuss their research interest in their own terms, exchange their ideas and opinions with peers or lay people, or take part in an online debate. Luzón (2009: 75) suggests that blogs are “gaining momentum as one of most versatile tools for online scholarly communication”. In this study, I aim at shedding some light into this novel phenomenon of blogs created and maintained by researchers. Below is a screen shot of a Web page that serves as an example of an academic research blog.

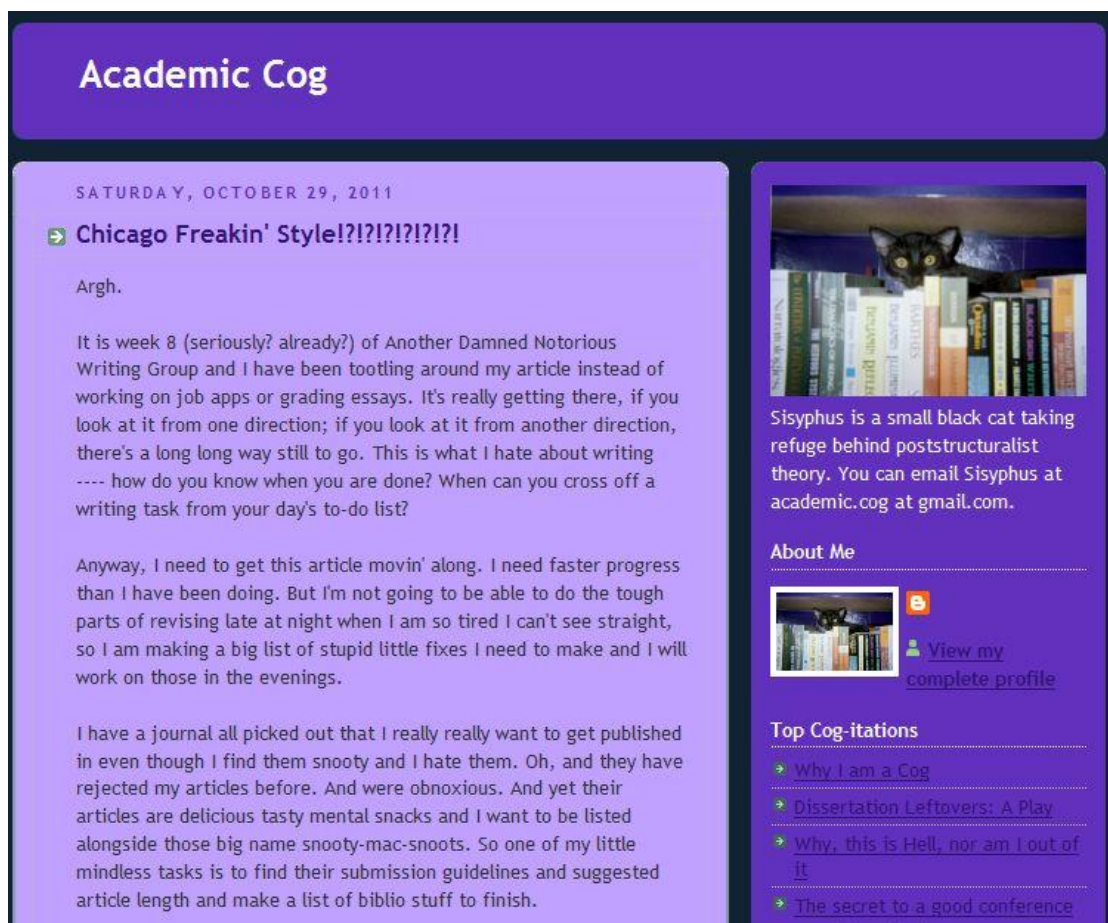


Figure 1: Academic Cog

There has been extensive research on different academic genres by various scholars (e.g. Bhatia 1993, Biber 2006, Hyland 2009 and Swales 1990), but, being a rather new phenomenon, academic research blogs have not been the focus of many genre analyses. Instead of attempting to form a general picture of academic research blogs as a genre, previous studies have often concentrated on some particular feature of the genre, for example on the reasons why academics blog (Walker 2006) or the practice of hyperlinking (Luzón 2009). The present study will try to fill the gap by providing an overall understanding of the genre of academic research blogs. This will be done by conducting a genre analysis, following the guidelines of Swales (1990) and Bhatia (1993), of randomly selected academic research blogs.

There are various justifications for carrying out a genre analysis of academic research blogs. First of all, the capabilities offered by computer network technology are reconfiguring the ways texts are produced and understood, giving an interesting setting for investigating the new genres of the World Wide Web. Second, being a very popular mode of

online writing, blogs can be considered an interesting online genre to analyse, because they highlight the affordances of hyperlinking and increased possibilities of multimodality not present in printed texts. What is more, blogging is a multi-faceted practice where seemingly personal texts are shared with the entire world in an informal style, which may run foul of the writing practices the blogging researchers have become accustomed to in the world of academia.

It is this collision of formal and informal, and public and personal that makes academic research blogs an interesting topic of research. What kind of text is formed in the cross-section of the immediate and free-form writing of blogs and the slow-paced and institutional writing of academic discourse? To find out about this, I decided to carry out a genre analysis of academic research blogs. To organise the thought and analysis process of the present study, I formulated the following research questions:

1. What are the generic features of academic research blogs?
2. How do academic blogs differ from blogs in general?
3. How does the language of academic research blogs differ from that of academic discourse in general?
4. Where do academic research blogs fit in the genre ecology of academic discourse?

The first research question concerning the generic features of academic research blogs is answered by conducting a genre analysis of the blogs, that is, by looking at the discourse communities that use and produce the genre texts, analysing the communicative purposes of the blogs, and determining their structural and linguistic features. The findings of the analysis are compared to the findings of a similar genre analysis on blogs in general by Herring, Scheidt, Bonus and Wright (2004) in order to find out if there is considerable variation between the generic features of academic research blogs and blogs in general. To see if the language of academic research blogs differs from that of academic prose in general, the findings of the analysis of the linguistic features prominent in academic research blogs will be compared to the findings of Biber's (1988) similar analysis. Based on the social, functional and structural elements of the academic research blogs attained from the analysis, the place of the blogs in the genre ecology of academic discourse will be discussed.

The material for this study consists of 30 randomly selected academic research blogs written in English. The blogs that were accepted as material had to fill the following criteria of being established blogs, that is, consisting of more than three blog entries, and being active blogs, which was determined by the date of the most recent entry. Furthermore,

the blogs had to be mainly text-based. Different kinds of academic blogs abound on the Internet, from student coursework blogs to university bookstore retail blogs. To narrow down the scope of the study, I decided to limit the focus on academic research blogs, that is, blogs that are written and maintained by university researchers or groups of researchers. In order to keep the corpus of the blog texts in a manageable size, only the entries on the first page of the individual blogs were included, thus making up a corpus of 259 entries and approximately 135 000 words.

The analysis of the data will be based on Swales's (1990) notions of what features constitute and determine a genre. These are the discourse community that uses and produces the genre text, the communicative purpose of the text, and the structural and linguistic features of the genre text. The more technical side of genre analysis, from surveying existing literature to examining the linguistic features of the text will follow the instructions of Bhatia (1993). After identifying and quantifying the discourse communities, communicative purposes and structural, functional and linguistic properties of academic research blogs I will be able to compare my findings to those of Biber's (1988) earlier study on academic prose, and to those of Herring *et al.* (2004) on genre analysis of blogs in general.

The theoretical framework for this study is formed of three wider topics. First, I will go through the focal points of genre theory, focussing on Swales's (1990: 45–58) definition of genre, Bhatia's (1993) notions of the goals of genre theory, and Miller's (1980) views on genre as social action. As one of the goals of this study is to see academic research blogs in relation to academic discourse as a whole, I will also discuss the main theories and studies in academic genres and academic discourse. These include, among others, the many works of Biber (1986, 1988, 2002a, 2002b, 2006) and Hyland (2001, 2002a, 2002b), and also Swales's (1990) studies on academic genres. Academic research blogs are discourse that is disseminated through networked computers, and therefore belong to the domain of computer-mediated communication (CMC), and this field of research is introduced by presenting the works of Herring (2001), Herring *et al.* (2004) Herring *et al.* (2005) and Androutsopoulos (2006) on the subject matter, before focussing on blog research, a specialised line of CMC especially relevant to the present study. A few studies on blog research have concentrated on academic blogs from various angles, and these studies are also briefly presented in the theory section of this study.

The aim of this study is to describe the functional and structural elements of academic research blogs through the method of genre analysis. The expected results of this research are the forming of an overall understanding of the underlying generic features of academic research blogs, and the determining of the characteristics that set academic

research blogs apart from blogs in general. In addition, this study sets out to find out the place of academic research blogs in the genre ecology of academic discourse. There are, however, some limitations to this study and what can be achieved with it. Being a relatively small-scale study with a limited data sample, this study cannot provide a clear-cut, all-embracing outlook on the genre of academic research blogs. What is more, Internet genres are in constant state of evolution, and what is a common practice today can be completely forgotten tomorrow. Therefore, the results and implications of the present study should be treated as suggestive, preliminary glances at the building bricks and defining features of the genre of academic research blogs of the present day.

It is important to analyse the vehicles used for enabling interaction that serves as a connecting bridge between fellow scholars around the world to discuss their research interests, and as a bridge to connect science and research with lay people – who, in many instances, are the sponsors of the research through tax paying. Academic blogs are a new and increasingly popular phenomenon in academic circles, and to fully understand their potential it is important to have an idea of their place in both the world of academia and the blogosphere. My study will try to shed some light into the questions of why and how academic research blogs are being written.

I will begin my study with presenting the theoretical background of the research. I will concentrate on the main issues on genre theory, academic discourse and computer-mediated communication, with special emphasis on blog research. After introducing the relevant background theories, I will move on to discussing the methodology used in the research. I will go through Bhatia's (1993) approach to conducting genre analysis, as this forms the backbone for the analysis in this study. Details on how the material for the present research was collected and analysed will also be dealt with thoroughly before turning onto the results of the analysis. In the section dedicated to the results and analysis, I will introduce the results of the genre analysis on academic research blogs. This chapter is divided into four sections according to the genre characteristics analysed, starting from presenting the results on the analysis of the discourse community, and moving through explaining the findings of the analysis of the communicative purposes of the blogs on to looking at the results of the structural and linguistic analyses. The findings of this research will be compared with the findings of relevant earlier studies where possible. After presenting the results and comparisons to earlier research, I will go through the implications of the findings in more detail in the discussion section.

2 THEORETICAL BACKGROUND

When analysing the genre of academic research blogs, one has to start building the theoretical framework from multiple points of view. First, it is essential that the notion of genre and the theory behind genre studies are taken into consideration, as it would be impossible to correctly conduct a genre analysis something without the theoretical knowledge of what constitutes a genre. Second, the term *academic research blog* points to the idea that the blogs contain academic language in one form or another, and therefore I will introduce the main concepts of academic discourse in order to be able to see what features of academic discourse are present in the discourse of academic research blogs. The affordances of networked computers, for example the affordance of hyperlinking texts to other texts to form nonlinear texts, offer new ways to produce discourse, and to be able to analyse the effect of computers and the Internet on the computer-mediated texts analysed in this study, I will first introduce the focal points of the study of computer-mediated communication.

2.1 GENRE THEORY

There is no universally accepted definition of genre, although many scholars have tried to come up with one, making genre theory a theoretical minefield where different definitions of and viewpoints on genre abound. Therefore, before embarking on an attempt at genre analysis, one has to make clear the theoretical viewpoint of genre on which the analysis will be based. The theoretical framework of this study will be synthesised from the viewpoint of Swales (1990) and Bhatia (1994), who look at genre in terms of consistency of communicative purposes, and that of Miller (1984), to whom the analysis of typifications of rhetorical action is at the core of genre studies. The theories of Swales (1990), Bhatia (1994) and Miller (1984) on genre were chosen for this study because they are in many cases the core theories that many of the later works on genre build on, and because they deal with genre on such a general level that they are easily adaptable to applied ends. Moreover, these theories have been used as the basis of most of the studies to which I will compare my findings, and having the same theoretical framework will make the comparisons more reliable.

2.1.1 *The concept of genre*

One of the best known genre theorists in linguistics is John Swales. In his monograph *Genre Analysis* (1990) Swales takes on the ambitious challenge of defining genre and making sense of

the fuzzy theory surrounding the concept of genre. Swales (1990: 45–58) argues that a genre must fill the following five criteria:

1. A genre is a class of communicative events.
2. The principal criterial feature that turns a collection of communicative events into a genre is some shared set of communicative purposes.
3. Exemplars or instances of genres vary in their prototypicality.
4. The rationale behind a genre establishes constraints on allowable contributions in terms of their content, positioning and form.
5. A discourse community's nomenclature for genres is an important source of insight.

A communicative event is one where language plays a pivotal, not merely incidental role. So, for example, a telephone conversation is considered a communicative event, whereas ice-skating is not. However, a communicative event does not mean just the language itself, but is consisted of the participants, discourse, the cultural and historical environment of the production and reception of the discourse. These communicative events must, in addition, occur often enough according to their prominence within the society so that they can be thought of as a genre in their own right. Papal Encyclicals occur very rarely but are a genre class, but a TV advertisement with a talking dog is not (Swales 1990: 45–46). Academic research blogs can easily be thought of as a class of communicative events, as language clearly plays a pivotal role in them, and while they are not as prominent within the society as Papal Encyclical, they do occur more often and often enough to constitute a class of communicative events.

The second criteria a class of communicative events has to fill to pass as a genre is to have some shared set of communicative purposes. So, Swales 1990: 46–49) takes the same stance as Miller (1984) in that it is the shared purpose, rather than similar form or content that constitutes a genre. For some genres, for example cooking recipes, identifying the shared set of communicative purposes is straightforward, but to some others it can be somewhat complicated. A good example of this comes from the academic world: the communicative purpose of students' written examinations can be hard to define, because in most cases the students are not writing essays to tell the reader about the rise and fall of the British Empire, but to show the reader, that is, the teacher, that they have learned the facts they are required to learn during the course. In order to find out if academic research blogs share a set of communicative purposes, I will analyse the communicative purpose of each blog

in the data set of this study and determine whether they form a set, and further, whether they form a class of communicative events.

The third criterion for a genre has to do with the properties and features which have to be present in all of the instances of a certain genre. Instead of a shared list of specific features, the membership of a certain genre seems to be based on quite loose inter-relationships. For instance, a lecture cannot be defined as an event where a teacher teaches students, because an event can still be defined as a lecture even if the teacher sits still without saying a word and students give their peers a presentation. A lecture with a teacher lecturing to the students is, however, more typical a lecture than a lecture consisting only of student presentations. So, the lecture with a lecturing teacher can be considered a prototype lecture, in the similar fashion as an apple is a prototype fruit instead of an olive. Swales (1990: 49–52) notes that whereas communicative purpose is the privileged property of a genre and therefore does not allow for much variation, content, form, structure and audience expectations are less important and thus vary according to the extent to which the instance is prototypical of a certain genre. By analysing and quantifying the structural and linguistic features of the academic research blogs in the study, I expect to find out the prototypical features of academic research blogs and how much variation there is in the prototypicality of the blogs.

The fourth criterion for genre is about the rationale behind the genre. According to Swales (1990: 52–54), the recognition of the shared set of purposes mentioned in the second criteria provides the rationale behind the genre, and this rationale brings about conventions according to the content, positioning and form, and constrains lexical and syntactic choices in the genre. Established members of a discourse community recognise and understand the rationale very well, whereas non-established members are not yet as good at perceiving the rationale. Being a rather novel and non-established genre, I expect to find a lot of variation in the content, positioning and form of academic research blogs.

The fifth and final criterion Swales (1990: 54–57) suggests puts emphasis on the discourse community's naming practice. As the fourth criteria suggests, established members have greater knowledge of the conventions of a genre than occasional members. These active expert members give genre names to those classes of communicative events they think are producing recurrent rhetorical action.

2.1.2 Main goals of genre theory

Along with Swales, Bhatia (2002a: 5) also considers genres forming of consistencies in communicative events, emphasising the amount to which the complexity of the world of discourse affects generic research. Instead of a proscriptive model, generic description should be viewed as a resource for “the knowledge of procedures, practices, and convention that make the text possible and relevant to a particular socio-rhetorical context” (*ibid.*). According to Bhatia, the main goals of genre theory are:

1. To represent and account for the seemingly chaotic realities of the world;
2. To understand and account for the private intentions of the author, in addition to socially recognised communicative purposes;
3. To understand how language is used and shaped by socio-critical environment; and,
4. To offer effective solutions to pedagogical and other applied linguistic problems.

To reach these four goals, it is essential, Bhatia (2002a: 8) argues, that one has understanding and knowledge of the following four perspectives on the universe of discourse. The first perspective is the real world perspective, in which discourse is dynamic and complex, because it is constantly evolving. This continual development is brought about by variation between different registers, disciplines, and cultures (Bhatia 2002a: 8–11). Another noteworthy factor for the complexity of discourse is the fact that genres tend to be an organised chaos similar to the star constellations in the sky. Bhatia (2002a: 8–10) explains how this chaos of genres is organised in colonies, or, constellations, of similar kind of genres which are systematically inter-related. More often than not, different genres form a complex network of inter-related genres, where an act or move in one genre results in and is followed by a certain appropriate generic response (Bazerman 1994: 96–97). Bazerman (*ibid.*) gives legal genres as an example of a web of genres that interact with each other. Correspondence, negotiations, appeals and court judgements are genres in their own right, but they are tightly knit to one another in the process of handling legal matters in a legally appropriate way. The constellations formed by academic genres have been the focus of several studies (see section 2.2.2), and one of the goals of this study is to find out how academic research blogs are related to other genres of academic discourse.

The second perspective Bhatia (2002a: 11–12) talks about is the socio-cognitive perspective. From this perspective, a genre theorist should be able to understand how expert members of a discourse community exploit generic conventions and rhetorical resources

across genres to achieve the desired communicative purpose. As an example Bhatia gives fundraising brochures, which are non-promotional and informational at their core, but where rhetorical resources of promotional genres are used to achieve the communicative purpose. As mentioned above, some communicative events serve a dual communicative purpose, which Bhatia calls mixed forms, such as academic course advertisements and book blurbs. Being a relatively new genre form, academic research blogs provide an interesting focus for studying what kind of generic conventions and rhetorical resources evolve in the early stages of the formation of a genre-text, and to see if they are used to serve a dual communicative purpose.

One crucial step in any kind of analysis of discourse as a genre is to choose the appropriate theoretical framework. Bhatia (2002a: 13–14) takes this as the third perspective of the world of universe, calling it the analytical perspective. These frameworks include corpus studies, textual analysis, genre studies and critical and ethnographic studies. The theoretical framework for the present study leans heavily on genre studies with a corpus-analytical approach to the analysis of the linguistic features of the genre-texts. Ethnographic approach to studying the genre of academic research blogs would surely have provided interesting results especially on the reasons for blogging and the effect blogging has on the academic blogger, but as the focus of this study is on the genre-texts and on what can be deduced from them, genre- and corpus-analytical approach was deemed more suitable for the present research.

The fourth perspective, the pedagogical perspective, deals with the pedagogical procedures and practices that are used to mould students to members of disciplinary communities. According to Bhatia (2002a: 14–15), knowledge of these procedures and practices is essential in making the analytical findings of generic research relevant to applied linguistic means, mostly for language learning (Bhatia 2002: 14-15). This perspective will not be taken into account in the genre analysis conducted in this study. A pedagogically oriented genre researcher could undoubtedly find a way to use this analysis to develop language learning, but coming up with ideas on how to use the findings of the genre analysis of academic research blogs for applied ends is outside the scope of this study.

2.1.3 Genre as social action

Genres have traditionally been defined according to the communicative purpose and recurring features of content and form in a text. These genres have then been used as simple typological taxonomic categories (Erickson 2000). In the 1980's, however, an idea was put

forward by Miller (1984) that the features of content and form of discourse which characterise a genre should not be considered as static conventions, but rather as vehicles for achieving the communicative purposes in a certain situation. Thus, the content and form are viewed as dynamic features which arise out of the technical, social and institutional forces which are present in the communicative situation (Freedman & Medway 1994: 8–9). In this approach to genre theory, genres are abstract conceptions, not something that exist empirically in the world. According to Miller (1984: 163), genres evolve, develop and decay, which brings the notion of evolution to genre theory. For example Bazerman (1988) has studied how the socio-cultural changes and other phenomena in the context of scientific knowledge have moulded and developed the genre of the research article, and, as blogs are one form of the rapidly evolving computer-mediated discourse, the diachronic perspective of the genre of academic research blogs is important to take into account.

Miller's (1984) influential and much cited article *Genre as a Social Action* proposes that genre research is useful in that understanding and knowledge of genre can be of help in identifying the ways people create, interpret and react to particular texts. She puts special emphasis on social and historical aspects of rhetoric in her analysis of genre. Miller (1984: 152) argues that "[a] classification of discourse will be rhetorically sound if it contributes to an understanding of how discourse works – that is, if it reflects the rhetorical experience of the people who create and interpret the discourse". In order to carry out this rhetorically sound classification of discourse Miller (1984: 155–158) suggests focusing on recurrent rhetorical situations, which are created from typifications of similar instances of situations. In conclusion, Miller (1984: 163) proposes a list of five features of her understanding of genre:

1. Genre refers to a conventional category of discourse based in large scale typification of rhetorical action; as action, it acquires meaning from situation and from the social context in which that situation arose.
2. As meaningful action, genre is interpretable by means of rules; genre rules occur at a relatively high level on a hierarchy of rules for symbolic interaction.
3. Genre is distinct from form: form is the more general term used at all levels of the hierarchy. Genre is a form at one particular level that is a fusion of lower level forms and characteristic substance.
4. Genre serves as the substance of forms at higher levels; as recurrent patterns of language use, genres help constitute the substance of our cultural life.

5. A genre is a rhetorical means for mediating private intentions and social exigence, it motivates by connecting the private with the public, the singular with the recurrent.

2.2 ACADEMIC GENRES

In order to find out where academic blogs are situated in the genre ecology of academic discourse, it is important to first discuss what academic discourse is. As the majority of genre studies in linguistics deal with discourse in academic and scientific settings, there is no shortage of relevant material. In the following section I will go through the selected works of Biber (2006,) Biber *et al.* (2002, 2004) and Hyland (2009) with regards to academic discourse in general, after which I will focus on Swales's (2004) notions of academic discourse as a genre. There are multiple reasons for why these studies are relevant to this study. First, these studies offer a clear idea on what is academic discourse and what are the characteristic features of it, which makes it possible to determine whether the discourse in academic research blogs can be considered a type of academic discourse, and to analyse which features of academic discourse are present in the academic research blogs. Second, the studies introduced below include research on corpora of academic genre texts, which are extremely relevant to the present study as a corpus will be compiled of the academic research blogs that serve as the data, and the findings of the corpus analysis will be compared to the earlier corpus findings on academic prose to find similarities or differences in the lexico-grammatical features of the texts. Finally, as one of the goals of this study is to find out where academic research blogs are situated in relation to other academic genres, I will go through Swales's notions of different ways to organise academic genres.

2.2.1 *Previous research on academic discourse*

Simply put, academic discourse refers to the way of using language that exists in the world of academia, enabling universities to go about teaching students and doing research. Hyland (2009) distinguishes four categories of academic discourse, the first of which are research discourses, that is, discourse that is used to convey information and knowledge in research settings between researchers. The second category is instructional discourse, the discourse used in the universities to teach students. Student discourses refer to the discourse that university students themselves use, and popular discourses are discourse used, for example, in science journalism to convey academic knowledge and information to the wider audience.

However, the role of academic discourse is not only that of conveying knowledge and information, but also shaping the social roles and relationships in the

universities and constructing the knowledge itself (Hyland 2009: 1–2). As Hyland (2009: 2) sums up, “the academy cannot be separated from its discourses and could not exist without them”. Discourse in the academy is the way new members are educated, academic allegiances are defined, collaboration and competence is carried out, and knowledge is created, and therefore the study of academic discourse can give information on social practices in the academic world (*ibid.*).

Most of the research in academic language has been conducted for applied, pedagogical ends (Flowerdew 2002: 2). In much of the research, the emphasis is on finding out the linguistic characteristics in different registers and genres so that students could be taught the differences between the specific kinds of discourses they have to master. In addition, most of the research on academic discourse is focussed on the linguistic features of written academic registers (Biber 2006: 6). Otherwise studies in academic discourse have been considerably diverse. For example, Biber, Conrad and Leech (2002) have compared the features of academic prose to that of fiction, conversation and newspaper coverage, Halliday and Martin (1993) have concentrated on the complex types of noun phrase structures that are typical of academic prose, and Hyland has an impressive slew of studies on a number of features in academic discourse, for example on self-mention in research articles (2001), on specificity and lexical bundles (2009), directives (2002a) and identity (2002b). In addition to these and other specialised linguistic features, like signalling topic, focus, anticipatory *it* or existential *there*, many studies have recently been centred around the topic of academic vocabulary (Biber 2006: 7). With these studies corpora of academic texts have been employed to elicit information on, for example, collocations in academic prose (Biber 2006: 7–8). Relevant to this study are especially the earlier corpus studies on academic language, as the material of the present study will be compared to the findings of earlier corpus studies on academic language. Next, I will discuss in more detail the work of Biber *et al.* (2002) on the *Longman Grammar of Spoken and Written English* (henceforth LGSWE), and Biber’s *University Language* (2006).

The LGSWE differs from most of the previous studies on academic prose in that it offers a comprehensive linguistic description of academic language, whereas most research focuses on one particular linguistic feature (Biber 2006: 13). The LGSWE compares academic prose to fiction, conversation and newspaper texts, and this comparison is based on the Longman Spoken and Written English Corpus, a corpus of about five million words from each genre (*ibid.*). The LGSWE describes not only the grammatical and lexical features in English but also the patterns of use for these features (Biber 2006: 13–14) and is therefore an excellent source for analysing the generic structures in English. Relevant to this study is the academic

prose part of the corpus, which consists of both academic research articles and academic books, containing 2.68 million and 2.65 million words, respectively. Texts from different academic disciplines are not separated, but the corpus represents academic prose as a general genre (Biber 2006: 14, Biber *et al.* 2002: 7–9). By comparing the frequency of different linguistic features, Biber *et al.* (2002: 7–9) found various features that were especially more common in academic prose than in other genres, and these can be considered characteristic of academic discourse. The most prominent of these features are the high frequencies of nouns and noun phrases, and adjectives and pronouns (*ibid.*, Biber 2006: 14– 18).

Whereas the LGSWE concentrated on distinguishing the genre of academic prose from other genres, Biber's later corpus study (2006) focuses on mapping the range of academic genres by providing linguistic description of different university registers. The data for the study is drawn from the TOEFL 2000 Spoken and Written Academic Language Corpus (T2K-SWAL Corpus). The T2K-SWAL corpus is a 2.7 million word corpus of the spoken and written registers that students come across in academic settings. It includes both academic registers, such as textbooks and class sessions, and institutional registers, for example course syllabi and service encounters (Biber *et al.* 2004: 7– 8). All in all, the corpus represents ten different university registers. The linguistic descriptions that were surveyed in Biber's study from each academic register covered by the corpus include the vocabulary distribution, lexical bundles, grammatical variation, expression of stance and multi-dimensional patterns of variation. The findings of the study show that different university registers have distinctively systematic patterns of use. Different linguistic features are distributed very systematically according to their associated communicative purposes in the world of academia (Biber 2006: 214). One of the most striking findings was that instead of factors such as different communicative purposes or degrees of interactivity, the most distinctive factor in determining the variation in linguistic patterns across university registers was whether the mode of discourse was written or spoken (Biber 2006: 213–218). In addition, what Biber (2006) found out was that student management registers, which was a general communicative purpose in the corpus, share many of the same linguistic characteristics, both in written and spoken modes of discourse (Biber 2006: 218–221). In comparison, there were hardly any shared linguistic characteristics in academic instruction registers across written and spoken modes, meaning that textbooks and classroom teaching employed very different linguistic features (Biber 2006: 221–223). Most spoken registers shared the same set of linguistic features, regardless of differences in audience and interactivity. So, the informational monologue of a lecturer had many of the same characteristics as a much more informal conversation between university students (Biber 2006: 223). What came up in Biber's (2006: 224–225) analysis of

lexical bundles, modal verbs, adverbial phrases and clauses, and complement clause constructions was that these features were often used in large amounts to express stance, particularly the personal expression of intentions, attitudes and evaluations of certainty, and interpersonal expression of directive language.

Biber's findings (2006) pose a very interesting question regarding the present study. Blogs are often written in such an informal style that the discourse can be thought to resemble spoken communication in many ways. Therefore there will be special emphasis in the linguistic analysis on the lexico-grammatical features that signal spoken and informal discourse.

2.2.2 Academic genres and their organisation

The universe of academic genres can be organised into different genre constellations in many ways. In *Research Genres* (2004: 12–25), Swales talks about hierarchies, genre chains, genre sets and genre networks. Genre hierarchies refer to the way in which members of a discourse community rank the research genres of their field in a certain order of value (Swales 2004: 12–13). For instance, Swales (1990) puts the empirical research article at the highest place in the research genre hierarchy, although in his later work he considers this a simplification of a complex interplay between academic genres (Swales 2004: 13). For example, for botanists, the most revered research genre is not the research article, but detailed descriptions of grouping of plants called monographs and flora, collected by an individual researcher over perhaps even decades. In the multidisciplinary automobile crash-safety research, on the other hand, the most central research genre is the conference presentation paper, a peer-reviewed, preprint of a conference presentation (Räsänen 1999). According to Swales (2004: 14–15), the most prominent factor accounting for this complexity in the hierarchical structures is the differences between disciplinary communities, and these differences were analysed in detail in Becher's (1989) influential study.

Becher (1989) divides the world of academic discourse into two groups, urban and rural collectivities. In urban research communities, for example in medicine, the knowledge is cumulative in nature, the people-to-problem ratio is high, and scientists usually work in groups, research is very competitive and therefore the pace of publication is fast. Hence, short and easily publishable research articles are the favoured mode of publishing. In rural groups, such as psychology or humanities, on the other hand, scientists usually work alone over a long time period, delving into complex problems. As the people-to-problem in rural areas of research is low, there is not much competition as to who is going to publish

one's findings first. Therefore the pace of publication is rather slow, and the central genre and favoured mode of publication is the scientific monograph (Becher 1989). What has to be taken into account, however, is that there are many research genres that do not seem to have a fixed place in many disciplinary hierarchies. For example, the hierarchical place for review article genre is in many disciplines undecided (Swales 2004: 16–18).

In hierarchies, genres are organised according to their perceived rankings. Research genres can also be organised chronologically in what Räsänen (1999) calls genre chains. A basic and simple example of a genre chain in academic discourse is an invitation to give a presentation at a conference, acceptance, possibly a preprint of the presentation, and the presentation itself (Swales 2004: 18). Genre chains can, however, be greatly more elaborate, as shown in Figure 1 below:

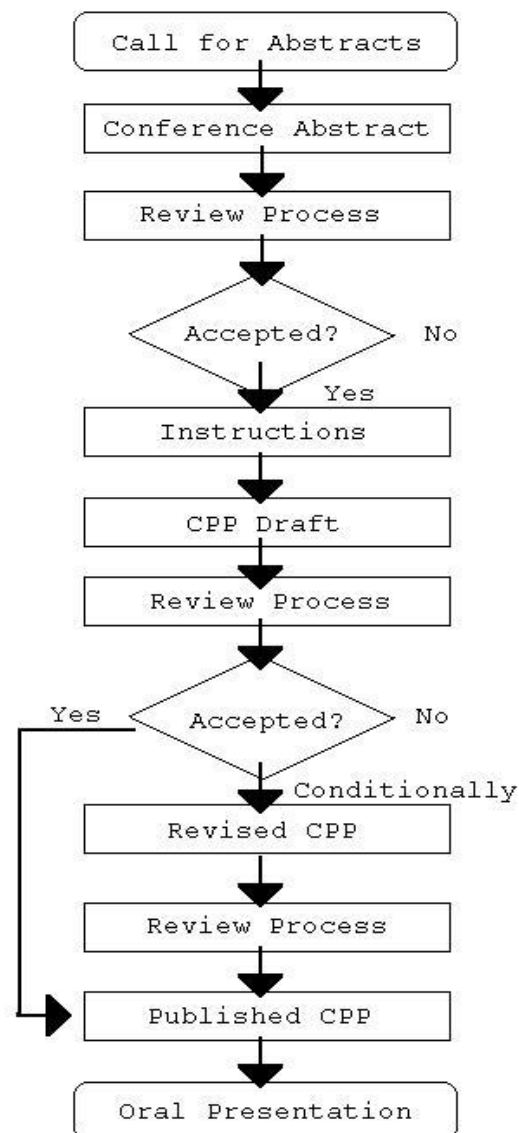


Figure 2: The conference proceedings paper genre chain (Räsänen 1999: 112)

Genre set, a concept coined by Devitt (1991), refers to the set of genres as individual encounters in their institutional practice. Devitt (1991: 339–340) defined thirteen genres that tax accountants write, and proposes that these thirteen genres form a set that reflects relations, activities and situations tax accountants encounter in their profession. Each of the genres in the set reflects a different rhetorical situation that the tax accountant comes by in their work (Devitt 339–340).

Lastly, Swales suggests looking at the genre ecology of the research world from the perspective of genre network. New genres, according to Todorov (1990) develop from earlier generic forms from already existing genres. This brings about what Devitt (1991) calls “generic intertextuality”. An example of making use of the rhetorical and linguistic features of an already existing genre is using speech elements in academic e-mails (Swales 2004: 21). These inter-connected genres then form a genre network, the genres of which are available for the certain discourse community (Swales 2004: 21–22).

In conclusion, the chaos of academic genres can be organised in many different ways from chains to networks. After identifying the generic features of the academic research blogs, I will discuss the place of the blogs in the genre organisation of academic discourse.

2.2.3 Linguistic features of academic discourse

I will now turn onto looking at lexico-grammatical features that are, according to previous studies, considered typical of academic language. These features are introduced below because their frequencies in the data are studied and analysed in the present study.

Based on Biber 1988, Biber *et al.* 2002, Conrad 1996 and Conrad 2001, Biber and Conrad (2009: 116–117) summarise the frequencies of the most prominent linguistic features in academic prose. According to the summary, the nominal features of nominalisations, prepositional phrases after nouns and attributive adjectives are extremely common in academic prose, while personal pronouns are rare. Of the verb characteristics, present tense is much more common than the rare past tense, and modal constructions are relatively uncommon, but passives cover about one fourth of all finite verbs. According to Biber and Conrad (*ibid.*), the circumstance adverbials of time and place are rare in academic prose, while linking adverbials are very common.

Whereas Biber and Conrad (2009) focussed on mapping all the grammatical features that are prominent in academic discourse, many other studies concentrate on the usage of certain grammatical constructions in academic language. For example, Hiltunen (2010) analysed the use of three grammatical constructions in research articles in four

different disciplines, and found out disciplinary differences between the soft sciences and hard sciences in the usage of declarative and interrogative content clauses and *as*-predicative constructions. In many other studies, systematic grammatical variation has been found between corresponding genre-texts from different disciplines (e.g. Charles 2003, Charles 2006a, Charles 2006b, Charles 2007) and between different genre-texts (e.g. research articles and book reviews) from the same disciplinary culture (e.g. Groom 2005). All this goes to show that while it is possible to see general trends on the prominent grammatical features of academic discourse, there is still considerable variation inside the whole scope of academic prose.

2.3 COMPUTER-MEDIATED COMMUNICATION AND WEB GENRES

The Internet has become mainstream and an everyday phenomenon in the Western Countries. The Internet makes it possible for people to communicate with each other regardless of their physical whereabouts. Linguists, cultural theorists and several other scholars have become interested in the communication that happens between human beings on the internet via the instrumentality of networked computers. The study of this is called computer-mediated communication, hereafter CMC, and it is a relatively new area of research.

In his article on sociolinguistics and computer-mediated conversation, Androutsopoulos (2006: 419–420) criticises the way media and even academic works on the subject have dealt with language use on the Internet as a non-standard register of language, as something that is incomprehensible to outsiders. This approach to CMC can be seen in the various names that have been suggested for language of the internet, such as netspeak, weblish and tech-speak (Thurlow, Lengel and Tomic 2004: 118). In the earlier studies on CMC, such as Crystal (2001), the main goal was to find general linguistic features that set CMC apart from other modes of discourse. According to Crystal (2001) and Herring (1996), these features are the use of emoticons and acronyms, the combining of written and spoken features, and the division between synchronous and asynchronous modes of digital communication (Androutsopoulos 2006: 420). However, these studies paid less attention, if any, to the social, contextual and technological factors in CMC, and computer-mediated communication was seen as a rather homogenous use of language (Androutsopoulos 2006: 420–421). Later on, more emphasis has been put to studying the sociolinguistic factors that the first researchers on CMC overlooked, such as the social situational factors like identity and community. The currently active areas of research in CMC revolve around the social, cultural and technological

factors that shape computer-mediated communication and break it into distinguishable genres (Herring 2001). Androutsopoulous (2006: 421) suggests that “rather than identifying e-mail, chat or weblogs as new genres per se, the question is how these communications technologies are locally appropriated to enact a variety of discourse genres”. Most of the research on Internet genres has tackled this question by looking at how different social and contextual factors affect and shape the studied CMC types, and this will be the viewpoint of the analysis in the present study, as well. In the next section, I will introduce some of these studies that have concentrated on analysing the genre ecologies in CMC in order to highlight the questions and problems related to genre analysis of computer-mediated communication, as these questions and problems will have to be dealt with in the current study, too.

2.3.1 Genre research on computer-mediated communication

With the growing amount and variety of communication in the World Wide Web, more and more interest has been taken in the identification of genres on the Web. Web-mediated documents have the affordances of being non-linear and multimodal, which poses new challenges when describing language use on the Internet. Some scholars have concentrated on creating a sound theoretical framework for analysing Web genres, while others have turned their focus on analysing the genres of specific types of Web documents, such as online encyclopaedias (Emigh & Herring: 2005) or personal home pages (Bates & Lu 1997, Chandler 1998, Dillon & Gushrowski 2000).

An example of the former type of Web genre research orientation is Askehave and Ellerup Nielsen's (2005) article where the authors suggest an upgraded version of the Swalesian genre model. The Web medium is an integral part of any Web genre, and therefore, according to the study, the navigational and hypertextual nature of Web documents has to be taken into account in the analysis of Web genres. Web documents should be understood in two modes: as traditional texts and as a medium which provides for the reader to navigate on the site. This double functionality of CMC causes problems in the traditional analysis of texts, because it makes the distinction between different texts very hazy. If, for example, a blog author refers in their blog entry to a study on a different Web page through a hypertext link, is the study behind the link to be considered in the analysis or not?

Shepherd and Watters (1998) have wrestled with the same problem of nonlinear online texts. Their article proposes that “while genres in other media can be characterised by the tuple, <content, form>, cybergenres are characterised by the triple, <content, form, functionality>” (Shepherd & Watters: 1998: 3). This functionality is an

affordance provided by the new medium of the Web, and gives rise to the formation of totally new kinds of genres. Already in 1992 Yates and Orlikowski suggested that communication in a new medium brings about not only adaptations of existing genres but also entirely novel genres. This has been proven to have happened with Web communication in a number of studies that focus on finding out where Web genres actually arise from. Shepherd and Watters (1998: 1–3) offer a taxonomy of Web genres, where cybergenres are first divided into two main subgenres, extant and novel genres. Extant subgenres are based on existing genres in other media, while novel subgenres are genres that are evolved in the new medium and do not have a counterpart in other media. In addition, they are wholly dependent on the new medium. Extant cybergenres are further divided into replicated and variant cybergenres, and novel cybergenres into emergent and spontaneous cybergenres, as can be seen in figure 2 below.

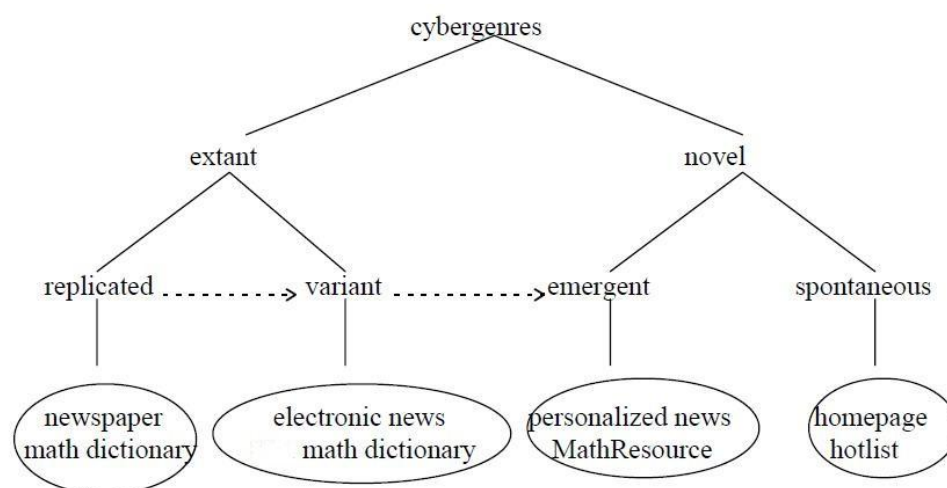


Figure 3: The evolution of cybergenres (Shepherd & Watters 1998: 2)

Replicated cybergenres, as the name suggests, are relatively faithful reproductions of the source genre they are based on, and the affordances offered by the new medium are little exploited. Variant subgenres, on the other hand, make to some extent use of the functionality brought about by the new medium. Variant genres are still firmly based on existing genres, but can have for instance hyperlinks or videos as opposed to just being plain text documents. Emergent cybergenres, a subtype of novel cybergenres, have evolved from existing genres so drastically that they have reached the point where they no longer share characteristics with the source genre, and can therefore be considered entirely new genres. Usually the evolutionary path of a Web genre is from replication through variant to emergent, and this evolution is caused by the progressive use of the functionalities afforded by the new medium.

Lastly, spontaneous cybergenres are genres that have no discernible counterpart in other media. An example of a spontaneous Web genre could be FAQs (frequently asked questions) or home pages (Shepherd & Watters 1998: 3).

Crowston and Williams (1997) mapped the genre ecology of the Web by randomly collecting 100 Web pages and then categorising them according to the type of genre they represented. The results showed that the data consisted of both examples of the reproduction and adaptation of genres. 80 of the 100 Web pages were classified as representing familiar genres, whereas 11 web pages were classified as new, but accepted genres. The rest, 9 web pages represented unknown genres in that their communicative purpose was unknown or they had very mixed features. In addition, Crowston and Williams (1997) encountered components of genre systems on the Web. In several Web pages examples of genres were embedded in other genres. Crowston and Williams (1997: 8) give an example of this from a page which was on the surface level an entry in an archive, but the entry included a letter which further included stories for a folklore collection. Each of these levels could be considered a new genre but they still carried the characteristics of the previous level. This embedding of genres is particularly relevant for the web environment, as text in electronic form is very easy to reuse, and because linking enables a single document to serve multiple purposes simultaneously.

Herring *et al.* (2004: 10–11) studied the place of blogs in the genre ecology of the Internet and concluded with the proposition that blogs serve as a bridge between standard and static Web pages and asynchronous, constantly updated CMC. It would be interesting to find out if the place of the genre of academic research blogs would differ from that of blogs in general, but the analysis of the place of the academic blogs in the genre ecology of the Internet is outside the scope of this study. However, the results of the present study show that academic blogs and blogs in general share many generic features, and therefore it can be hypothesised that their place in the genre ecology of the Web is roughly the same.

2.3.2 Blogs and blog research

The word *blog* comes from *weblog*, a log published on the Web. Maintaining a diary publicly accessible through a network has been technically possible since 1969, but according to Blood (2000), there were only 23 blogs known on the entire Web in 1999. This soon changed drastically when software designers saw a market niche and started launching different kinds of blogging software that made publishing blogs quick and easy (Blood 2000). Maintaining a

blog no longer required knowledge of coding HTML, and this lowered the threshold for more and more people to start publishing their daily thoughts to the world of WWW. Growth in the number of blogs has been rapid for the past 12 years, and as of 2011, there were more than 156 million public blogs (Wikipedia: Blog).

But a blog is not just a diary published online. If it was, there would be little point in analysing it as a distinct genre. Herring *et al.* define blogs as “Web-based journals in which entries are displayed in reverse chronological sequence” (Herring *et al.* 2005: 1) and the main point to focus on is the fact they are *Web-based* journals. The affordances of the Web create distinct differences between blogs and their source genres, journals. The Web makes it possible to make hyperlinks between documents, and it is this hypertextual and interconnected nature which makes blogs an interesting topic of study from many points of view. Bloggers can link to other bloggers, or refer to them in their posts, or post comments on other bloggers’ blogs, thus making the universe of blogs an interconnected system, often referred to as the “blogosphere” (Herring, Scheidt, Bonus and Wright 2005: 1). On the other hand, bloggers can provide links to any other Web site on WWW in their blogs, or even embed images or videos from other websites in the entries on their blogs.

Why then do people blog? This is an essential question in studying blogs as a genre, as the aim and purpose of producing discourse is a factor that has to be taken into account when determining the genre the discourse represents, and therefore there has been extensive interest in finding out the reasons for blogging. Some blogging scholars have done autoethnographical studies on their own motivations for blogging, some have resorted to interview studies, while others have examined blog entries to define the aims of the blogger. In their autoethnographic study, Davies and Merchant (2007) analyse their own reasons for blogging, which include, for instance, the developing of social networks, the improving of self-representation and textual practices. Hsu and Lin (2008), on the other hand, interviewed 212 blog authors and found out that the underlying motivations for blogging were manifold, from pure enjoyment through knowledge sharing to social factors.

Relevant to this study is Herring *et al.*’s (2004) groundbreaking genre analysis of blogs, which was the pioneering study to systematically describe the characteristics of blogs. Herring *et al.* collected 203 random blogs and analysed from the data blog author characteristics, the purpose of the blog, temporal measures and structural characteristics of the blogs. The results of the study shed light into the question of the origin of the blog genre, and also its place in the genre ecology of the Internet. According to the study, blogs have both elements that have developed out of previous Web genres, such as personal home pages or hotlists, and elements from previous offline genres, for example from diaries, editorials and

newsletters, thus forming a hybrid and evolving genre. In the study, blogs are found to bridge the technological gap between the relatively static standard Web pages and constantly updated, asynchronous computer-mediated communication.

2.3.3 Academic blogs

Academic blog is a term that includes numerous different genres from political blogs to pure research blogs (Walker 2006). Pure research blogs – also referred to as science blogs – have become an increasingly popular way of disseminating scientific research. Many scholars, science laboratories and research groups blog to openly discuss their research, and there are several scientific journal sites that host a collection of science blogs, such as the site of Nature Publishing Group. Luzón (2008: 76) suggests that the emergence and growth of academic blogging has its roots in the emergence of Web 2.0 and Science 2.0, which are both based on the idea of collaborative information sharing and knowledge construction and open access to intellectual engagement and research results and theories. In the past few years, several scholars have studied academic blogs from various angles. Some studies focus on analysing the disciplinary communities that blogs create (Efimova & Hendrick 2005), while others concentrate on analysing the linguistic and communicative features of academic blogs (Stuart 2006, Efimova & de Moor 2005). In addition, several studies have been conducted to analyse the uses of blogs in research and academic circles (Davies & Merchant 2007, Halavais 2006, Mortensen & Walker 2002). In the following section, I will briefly introduce a few previous studies on academic blogs that are relevant to the present study. These include a linguistic analysis of academic blogs by Stuart (2006), Luzón's research on academic hyperwriting (2009), and studies on using blogs as a research tool by Mortensen and Walker (2002), Walker (2006) and Kjellberg (2010). These studies are considered relevant because they offer the methodological backbone for the study, as the methods for applying genre analysis to analysing blogs in the present study are my adaptations of the methods used in the se studies.

Stuart (2006: 391–392) analysed the communicative purposes of 496 blogs within academic organisations, and came up with a list of 19 categories ranging from promoting the university to blogs as ePortfolio. He then further classified the communicative purposes into four more general categories, which, according to Stuart, can be considered to form “a subset of genres within the larger system of academic and educational genres which are intertextually and interdiscursively linked within the boundaries of academic and educational settings”. Below is a table representing this genre subset:

Table 1: Uses of blogs in academic institutions (Stuart 2006: 392)

Author	Student	Teacher
Researching	PhD / Masters Weblogs	Academic announcements, critique, debate
Learning / Teaching	Student weblog as coursework Student group discussion tool	Class weblog Publication of texts, links or commentary to seed discussion

As the focus of this study is on academic research blogs, educational academic blogs are outside the scope of this study and the rest of this chapter will be dedicated to delving deeper into the realms of blogs used for researching purposes.

According to Blood (2002: 59), “personal detail is not necessary [...] but every good weblog has a point of view”. Therefore, blogs tend to be openly evaluative and critical in nature, which contradicts the underlying idea of scientific inquiry and basing one’s arguments on scientifically proven facts in the world of academia. Stuart (2006: 392) suggests that this discrepancy between publishing one’s thoughts and ideas in a blog and publishing the results of one’s scientific inquiry in a research article may account for the fact that there are relatively fewer blogs in the hard science disciplines than in the soft sciences, where the research tends to be more based on subjective interpretation.

Stuart (2006: 393) sees that formal citation practice is one of the defining characteristics of academic writing that is also present in academic blogs. References in academic blogs seem to be random links to other sources, but on closer inspection they are usually found to be explicit references to other texts, blogs or web pages on which the blog author’s arguments are based. Furthermore, hyperlinks have been found to have numerous other functions in academic blogs. In his article on scholarly hyperwriting, Luzón (2009: 75) analysed the types of links in 15 academic blogs, and found out that blogging scholars used links “to seek their place in a disciplinary community, to engage in hypertext conversations for collaborative construction of knowledge, to organise information in the blog, to publicise their research, to enhance the blog’s visibility, and to optimise blog entries and the blog itself”.

Stuart (2006) studied the formal linguistic features of 39 academic blogs in order to compare the language of academic blogs to that of academic language in general. Stuart analysed 26 different linguistic features that are descriptive of academic language, and compared the findings of his study to Biber’s (1988) earlier study on academic prose. The results showed that in most of the features there were very little difference between the frequencies of use in academic prose and academic blogs. However, there was marked

variation in the use of first and second person pronouns. Perhaps because of the diary-like form and the conversational nature of academic blogs, the use of first and second person pronouns was considerably more frequent in academic blogs than in academic prose. Stuart's study (2006), however, focused in its genre analysis mostly on the quantitative linguistic analysis of the features descriptive of academic discourse. The present study takes a wider approach to genre analysis, focusing not only on the analysis of the linguistic features of the texts, but also on the discourse community that uses and produces the texts, and on the communicative purposes and structural features of the blogs. Moreover, Stuart's study was conducted in 2006, and it will be interesting to see if the language of academic research blogs has changed in the six years that separate these studies.

Mortensen and Walker (2002) both started to keep a blog to "keep their focus online" while working on their PhDs, but they soon had developed into a hybrid of a diary, storage space, academic publishing and a place for academic discourse. In their autoethnographic study they turned their focus on what writing blogs do to their academic thinking. What the computer game researchers Mortensen and Walker (2002) found out in the study was that writing a blog was a great asset for the online research they were conducting. This claim is validated by their argument that "rather than distancing ourselves and permitting an escape from the object of research, the blog lives within the same frame as the computer games and the electronic narratives we study, keeping us close to the technology, the relevant formal as well as informal discourse and the objects themselves" (Mortensen and Walker 2002: 273).

Four years later, Walker (2006) returned to analysing her own blogging and research blogging in general. She divides academic blogs to three main categories. The first is *public intellectuals*, a category which contains blogs that are used as a platform for political debate and discussion about current political events. The second category is *research logs*, which means blogs that are written as a record of the research that is being conducted by the blog author. The third category *Pseudonymous blogs about academic life* contains blogs that are less serious with topics that are more suitable for coffee breaks than a conference (Walker 2006: 4–5). Kjellberg (2010) also wanted to map the reasons for and functions of academic blogging, and conducted an interview study, asking blogging scholars about the functions their blogging serves for their work as a researcher. The results of the survey showed that blogging is motivated by several factors, mostly by the possibility to share knowledge, stay connected to the disciplinary community, and serving as a creative catalyst for writing. Below is a table which shows Kjellberg's view on the interplay between function and motivation for blogging in the academia.

Table 2: The interplay between function of and motivation for blogs use in a scholarly context (Kjellberg 2010)

Function	Audience	Motivation
Disseminating content	Others	Sharing
Expressing opinions		
Writing	Self	Room for creativity
Keeping up-to-date and remembering		
Interacting	Self and others	Feeling connected
Creating relationships		

Unlike the study of Kjellberg (2010), the present study does not have an ethnographic angle on the analysis, and therefore the reasons and motivations for blogging cannot be analysed. However, the results of the study show similarities in the functions and audience of blogs between the study of Kjellberg (2010) and the present study, and therefore it can be speculated that blogging is probably motivated by similar reasons in both studies.

3 MATERIAL AND METHODS

In order to map the generic features of academic blogs, a genre analysis of random academic research blogs was carried out. In the following subsections I will introduce the theoretical basis for the analysis of academic blogs that will be used in the present study. I will conduct the analysis of my data by following Bhatia's (1993) approach to genre analysis, and this approach will be covered in the first subsection. In the second section of this chapter I will go through the way the material for this study was collected, and in the third subsection I will explain how the data was analysed.

3.1 ANALYSING UNFAMILIAR GENRES

According to Bhatia (2002a: 6), genre analysis means "investigating instances of conventionalised or institutionalised textual artefacts in the context of specific institutional and disciplinary practices, procedures and cultures in order to understand how members of specific discourse communities construct, interpret and use these genres to achieve their community goals and why they write them the way they do". Bhatia (1993, 2002b) suggests a detailed framework for analysing unfamiliar genres. The framework consists of seven steps, which will be discussed below.

The first step of the framework for investigating any genre is *placing the given genre-text in a situational context*. This is largely done intuitively, basing the analysis on one's prior experience and knowledge of the world, and looking at the internal clues in the text. The ability to do this initial analysis is greater in those people who are a part of the discourse community which produces the genre-text, because they have greater background knowledge of the community and communicative conventions associated with it (Bhatia 1993: 22). So, in the present study, the initial placing of the genre-texts in a situational context is easier when one has basic knowledge of the academic communities and their writing conventions, as well as Internet language, computer-mediated communication and blogs as a genre.

The next step in analysing an unfamiliar genre is *surveying existing literature* (Bhatia 1993: 22–23). In order to refine the initial situational analysis, the genre analyst has to gain more knowledge of the text-producing speech community and its conventions. Bhatia (*ibid.*) suggests reading literature on, among other things, prior linguistic analyses of the genre in question, theories, tools or methods of genre analysis relevant to the study, and guide books and manuals relevant to the speech community. For this study, the step of surveying existing literature was carried out by looking into the literature on analysing various Web

genres and academic genres, and skimming through guides on how to be a publishing researcher, and also guides on how to keep a blog.

After initially placing the text in a situational context and surveying existing literature, the genre analyst continues with *refining the situational/contextual analysis* (Bhatia 1993: 23). The refined analysis is achieved by looking more closely at four different aspects of the text. The first aspect to look at is the speaker or writer of the text, its audience and the relationship they have with the author. Secondly, one has to analyse the socio-cultural and occupational aspects of the discourse community, followed by the identification of the network of texts and linguistic traditions that surround and form the background to the text. Finally, a genre analyst should go about identifying the reality the text attempts to represent and the relationship between the text and that reality. The analysis of the situational and contextual features of the genre-texts form the main part of the analysis in this thesis, and they are dealt with in chapters 4.1, 4.2 and 4.3 in the results and analysis section.

The fourth step in Bhatia's (1993: 23–24) view of genre analysis is *selecting corpus*. It is crucial to any study that there is just the right kind of data to be analysed. To select an appropriate and representative corpus for analysing unfamiliar genres one has to define the studied genre well enough for it to be distinguishable from related genres. This definition can be based on some distinctive textual features of the text, its communicative purposes, the situational contexts the text is produced, or a combination of these factors. In addition, Bhatia (*ibid.*) suggests making sure the criteria for a text to belong to a specific genre are clearly stated. Finally, the amount of data has to be decided on. The corpus can consist of one single typical text which is then thoroughly scrutinised, or the corpus can be a random collection of a few texts for exploratory investigation. There can also be a very large corpus for statistical analysis to look at a few clearly specified features. In this study, the selection of texts to be analysed is mainly based on their structural features and communicative purposes. The texts were identified as blogs according to their structure, and academic research blogs according to their communicative purposes of discussing academic and scientific topics. The corpus in the present study consists of 30 randomly selected blogs, and the size allows for both exploratory investigation and a small-scale statistical analysis through easily identified indicators.

The corpus selection is followed by *studying the institutional context* (Bhatia 1993: 24). This means analysing the way the genre is used and looking at the often implicitly understood and unconsciously followed rules and conventions that govern the language use in the institutional setting of the text. Relevant to this study is, for example, studying the citation

practices in academic writing and to see whether the formal citing practice is also evident in academic blog writing.

The penultimate step in analysing genre is *linguistic analysis*, which is done on one or more of the following levels of linguistic realisation:

Level 1: Analysis of lexico-grammatical features

Level 2: Analysis of text-patterning or textualisation

Level 3: Structural interpretation of the text-genre

The first level, *analysis of lexico-grammatical features*, usually involves a large-scale statistical corpus analysis of a sample of texts that is representative of the variety in question (Bhatia 1993: 24–26). On this level, the genre-text is usually studied by analysing the specific features that are often used in the variety the text belongs to. Studies of frequency of syntactic properties, such as uses of tenses or clause structures are examples of analyses of lexical-grammatical features. This level of analysis has been criticised for the fact that it mostly provides information on the surface features of the genre-text and does not give insight as to how these features work for accomplishing the communicative purpose of the text (*ibid.*).

This is where the *analysis of text-patterning or textualisation* comes into play. This form of linguistic analysis focuses on how various aspects of language are assigned restricted values by the members of the discourse community. For example, Swales (1974), studied the function of past-participles in pre- and post-modifying noun phrase positions in chemistry textbooks, and came to the conclusion that pre-modifying *en*-participles (e.g. *given*) textualised two different aspects in chemistry texts depending on whether they occur in contexts where the author of the text is generalising or exemplifying. In the contexts of exemplifying, the function of *en*-participles was to signal that the common convention of attribution and specificity in science texts was temporarily suspended in order to avoid irrelevant enquiries about the details of the experiment. When the author was making generalisations, the pre-modifying *en*-participles were signalling definiteness without commitment to this specificity. So, analysing the lexico-grammatical features of a genre-text may not yield interesting results, but looking at what the studied features textualise can lead to more insightful analysis of the genre (Bhatia 1993: 26–29).

The third level of linguistic analysis of a genre, *structural interpretation of the text-genre*, deals with the cognitive aspects of language organisation. In several studies, established members of a discourse community have been noted to be consistent in the way they organise information when producing a genre-text. One of these studies is that of

Swales's (1990) analysis on the structural organisation of information in scientific article introductions. What Swales found out was that there were remarkable similarities in the article introduction organisation between the researchers, regardless of their topic disciplines. Swales suggests the following cognitive move-structure for academic research article introductions:

Move 1: Establishing the research field

Move 2: Summarising previous research

Move 3: Preparing for present research

Move 4: Introducing the present research

So, these four rhetorical moves are used to accomplish the communicative purpose of the article introduction, and they form the typical cognitive structure for the research article introduction genre (Bhatia 1993: 29–34).

Due to resource constraints, the linguistic analysis in the present study will be done on level 1, that is, as an analysis of lexico-grammatical features. No part-of-speech tags were used in the corpus that provides the material for this study, so the focus of the level 1 analysis of lexico-grammatical features is mostly pure lexical analysis, leaving the analysis of most of the grammatical features out of the scope of the study. As the linguistic analysis is based on statistical analysis of a corpus of 135 000 words, there will be no analysis of the move structure of each blog entry, as it would require a detailed, qualitative analysis of the whole corpus.

The seventh and last step of Bhatia's (1993: 34–36) analysis of an unfamiliar genre is acquiring *specialist information in genre analysis*. In this stage, the genre analyst has gathered the findings of the analysis and double checks if the findings are correct in a specialist informant's opinion. A specialist informant can be anyone who is a practising member in the discourse community that uses the genre that is being analysed. According to Bhatia (1993: 34), "specialist reaction confirms [the genre analyst's] findings, brings validity to his insights and adds psychological reality to his analysis". In Bhatia's view, specialist information is an important part of any genre analysis, if one wants it to include relevant explanation in the analysis. However, due to time and other resource constraints, this genre analysis does not make use of a specialist informant's perceptions of the reliability and validity of the findings.

3.2 DATA COLLECTION

This study is based on an analysis of 30 randomly chosen academic blogs collected from November through December 2011. Initially, 502 blogs were collected from 86 British, Irish, Australian, New Zealandic, Canadian and American university Web sites and sites that host academic research blogs, including <http://scienceblogs.com/>, <http://researchblogging.org/> and <http://blogs.nature.com/>. In addition, academic blogs were searched using the search engine Google. To create a coherent corpus suitable for the present study, all non-English blogs and blogs that did not contain significant amount of text, such as photo and audio blogs, were excluded. Also discarded were academic blogs that were clearly written by non-researching graduate students, and blogs that had fewer than three entries or had not been updated in two months, as the focus of the study was limited to active and established academic research blogs. These deductions cut the number of blogs to 185. In the next phase, 30 blogs were selected randomly from these 185 blogs. In order to keep the corpus in a reasonable size, only the entries on the first page of each blog were examined. Thus, the number of entries per blog in the final corpus ranged from 5 to 10.

While including them in the qualitative analysis of the discourse communities, communicative purposes and structural features, boilerplate texts, such as headers, footers, navigation information and other content peripheral to the main content were left outside the scope of the quantitative linguistic analysis. Therefore, the text bodies of each entry in each blog were copied to pure text file format, *allblogs.txt*, ignoring the boilerplate content. As a result, a text corpus of 135 000 words was compiled from the blog entries.

As for the legal and ethical issues concerning the study, there seems to be no other problems except for that of whether the authors of the blogs should be asked permission to, or at least be informed of, a study being conducted of their writings. When it comes to intellectual property rights, Finnish copyright law permits a private person of using other people's online texts to compile a corpus for their own use. Therefore, there is no legal reason for contacting the blog authors, and considering that the blogs are freely available on the web for anyone who stumbles upon them, it was deemed unnecessary to inform the blog authors of this study.

3.3 DATA ANALYSIS

The analysis in the present study is divided into four sections according to what features of a text are relevant in genre analysis. These are the analysis of discourse communities,

communicative purpose and the formal features of structure and language. The data for the analysis of discourse communities, communicative purpose and structural features was obtained through content analysis of the blogs as a whole, including boilerplate texts and comment sections. The linguistic features were analysed through a statistical corpus study of the text bodies of each blog entry.

In order to situate the genre of academic research blogs in a discourse community, I analysed the demographic characteristics of the blog authors. These include the sex, country, title and field of research of the author, as well as the number of authors in each blog and whether they use their real name or a pseudonym. In addition, audience and reader characteristics were determined. These characteristics were collected by analysing the blog entries according to what kind of audience the author has intended the texts for, and looking at the comments section of the entries to see the characteristics of the readers. Also analysed were the amount and form of interaction between the blog author and the audience. This was done by determining if the authors refer to their audience in the blog entries and if there is any kind of dialogue between the author and the audience in the comments sections of the blogs.

Next, the overall communicative purposes of the blogs were coded. Each blog was assigned to one of the following four categories of communicative purposes according to what categories were most strongly represented in the blog: *invisible college* (i. e. blogs written for fellow academics in the field on field-related topics), *diary about academic life* (i.e. blogs about the academic life and work of the researcher), *popularising science* (i.e. blogs on field-related topics written with lay people in mind) and *soap box* (i.e. blogs where the academic blogger shares their opinions on controversial, field-related topics. These categories were synthesised from earlier similar studies of Herring *et al.* (2004), Halavais (2006), Walker (2006) and Kjellberg (2010). Below is an example of the analysis of the communicative purposes of the blogs.

Table 3: Example of the analysis of communicative purposes

Purpose	Respectful Insolence	Profgrrrrl	Aetiology	All That Matters
Invisible college				X
Diary About Academic Life		X		
Popularising Science			X	
Soap Box	X			

The structural analysis of the blogs was carried out by selecting appropriate structural features to be studied and counting their occurrence in the blogs. The selection of the features was based on earlier studies on Web genres (e.g. Herring *et al.* 2004). These features include entry length, images, videos, advertisements, software used, comments section, about page, archives and links. In addition to counting the number of links in each blog, the links were divided into four categories according to what kind of resource they were linking to. The link categories were formed based on the analysis of the most common functions of the links in the data. These categories are *reference*, *self-reference*, *explanation* and *entertainment*. Temporal features of the blogs were also measured, such as the frequency of posting and the age of the blog. All the analysed structural features are listed below:

- About page
- Advertisements
- Age of blog
- Archives
- Average entry length
- Badges
- Blogging software
- Blogrolls
- Calendar
- Comments per entry
- Date
- Feeds
- Frequency of posting
- Images
- Links per entry
- Links: entertainment
- Links: explanation
- Links: reference
- Links: self-reference
- Entries on the first page
- Tags
- Videos

The final stage of the genre analysis of academic research blogs was the linguistic analysis. The linguistic analysis was carried out on the lexico-grammatical level, and the analysis was limited as the words in the corpus were not tagged according to the parts of speech they represented. Therefore, the linguistic data could be collected with simple commands using standard GNU/Linux corpus tools. For example, first person pronouns were quantified with the following command line:

```
cat allblogs.txt | tr '[:upper:]' '[:lower:]' |
egrep -wio 'I|we|me|us|my|our|mine|ours|myself|ourselves' |
sort -nr | wc -l
```

For some features, simple calculations of word frequencies were not sufficient, and they had to be analysed in their context. For example, when counting the frequency of the emphatic *do + verb* one had to look at the context where the word *do* appeared to determine whether it was followed by a verb to see if it was used as an emphatic. If one would only count the instances of the word *do* one would get wrong results, as other constructions, like questions, would also be counted. So, it was necessary to go through all the instances of the word *do* in

their context to see if it was followed by a verb. In order to see the words in their context, a following command was formed:

```
cat allblogs.txt | tr '\n' ' ' |  
egrep -io '[:print:]{30}do [:print:]{30}'
```

This command produced a concordance list of the word *do* with 30 characters before and after it. Below is an example of the concordance list formed by the command:

a private activity. But if we	do	decide to share our data, how
no reason why I would want to	do	that as an individual, but for
ttt much exhausted what I can	do	that as an individual, but for
a little bit longer, though I	do	cover a few of the topics I ta
ts that followed, I decided to	do	a thought experiment. Imagine
the brain. To begin with, how	do	we expect our best games to
psychophysiological response.	Do	not change too many variables
they might want to stay on to	do	a Masters programme or seek

The feature frequencies were then normalised and analysed and compared to the results of earlier studies.

4 RESULTS AND ANALYSIS

In this section, I will present the results of this study. The findings of the research are shown in quantitative summaries, and these summaries are followed by more in-depth analysis and interpretation of the results. I will begin this section with presenting the results of the analysis of the discourse communities that use and produce academic research blogs. In the second section of this chapter I will look at what the results tell about the communicative purposes of the blogs. The final two sections are dedicated to the formal features of the blogs, that is, their structural and linguistic characteristics of academic research blogs.

4.1 DISCOURSE COMMUNITIES

To situate the genre of academic research blogs within a discourse community, blog author characteristics, audience and reader characteristics, and the interaction between the author and the readers were analysed. The following three subsections present the results of the analysis of these features of academic research blogs.

4.1.1 *Blog author characteristics*

One important task when analysing the genre of any text is to look at the discourse community that produces the text. The table below illustrates the general characteristics of the blog authors in the present study.

Table 4: General blog author characteristics

Characteristic	Frequency	Percentage
One author	26	86.7
Male	18.5	61.7
Located in USA	19	63.3
Full name	24	80.0
Graphical representation	20	66.7

It is apparent from this table that the majority (26 out of 30) of the blogs in the study were created and maintained by a single individual. The prevalent gender of the blog authors was male with 18.5 bloggers, compared to the amount of 9.5 of the blogs being maintained by female bloggers. Even though the blogs were collected from various English-speaking universities and blog directories around the world, the majority of the blogs (19 out of 30) was

written by residents of the United States. Seven of the blogs were written by bloggers living in the U.K. and the remaining four blogs were from Sweden with two instances, and Australia and Italy, both with one instance. According to the data, authors of academic research blogs do not tend to hide behind pseudonyms, with 24 blog authors giving their full name in their blogs. Displaying a graphical representation of the author was rather common, with 20 bloggers providing a photo of themselves in the blog.

The data on general blog author characteristics provide interesting differences when compared to the findings of the genre analysis of blogs in general by Herring *et al.* (2004). The demographical characteristics of gender and geographical location follow the results of Herring *et al.* with approximately similar percentages of male and female authors and the prevalence of U.S. bloggers. However, very striking differences arise when looking at how much information bloggers give about themselves. In the study of Herring *et al.*, only 31.4% of the blogs displayed the blog author's full real name, whereas the percentage is considerably higher in this study (80.0%). In addition, Herring *et al.* (2004: 5–6) found that only 17.5% of blog authors provided any graphical representation of themselves, which is a much lower percentage than the 66.7 per cent of the research bloggers in the present study showing a real or drawn image of themselves. The age of the bloggers was not taken into account when looking at the characteristics of blog authors, because most of the bloggers did not share any information on their age, and therefore there was not enough data on the age of the bloggers to allow for any analysis or generalisations.

For the academic system to work, it is necessary to have a clear hierarchical system within the university staff (Walker 2006: 2–4). To find out if this hierarchical system is displayed in academic blogs, the blogs were coded for whether they provided information on the author's title and university rank. Table 6 below presents the academic titles that the blog authors displayed in their blogs.

Table 5: Blog author title

Title	Frequency	Percentage
Postdoctoral researcher	10	33.3
Professor	6	20.0
Doctor	5	16.7
PhD student	4	13.3
Associate professor	2	6.7
N/A	2	6.7

The most common title in the data set was *postdoctoral researcher*, with one third of the blog authors defining themselves as 'postdocs'. Titles *professor*, *doctor* and *PhD student* were distributed quite evenly, with six, five and four per cent, respectively. Two bloggers defined themselves as *associate professors*. What is notable is that only two blog authors had decided not to reveal their university rank.

So, a staggering 93.3% of the blog authors displayed their title. This is a very high frequency considering that Herring *et al.* (2004: 5) found that only 55% of blogs in general mention the blog author's occupation. It seems that the importance of hierarchy in the academia is also present in the academic blogosphere, and it is manifested by the common habit of mentioning one's academic rank in one's blog.

As mentioned earlier in the theory section of this study, Stuart (2006: 392) proposes that hard sciences tend to produce fewer blogs than soft ones. In order to find out if this was the case in the material collected for this study, the blogs were categorised according to the field of science the blog belongs to. Table 7 below shows the summary of different fields represented in the data set.

Table 6: Field of science

Field of Science	Frequency	Percentage
Medicinal studies	4	13.3
Physics	4	13.3
Mathematical studies	4	13.3
History	3	10.0
Law and political science	3	10.0
English	3	10.0
Geo studies	3	10.0
Culture studies	2	6.7
Biological studies	2	6.7
Other	2	6.7

From the table above we can see that the academic blogs analysed in this study represent different fields of science very evenly. Most common were *medicinal studies* (in more detail, medicine, biomedical engineering, epidemiology and psychiatry), *physics* (physics and astrophysics), and *mathematical studies*, all with four blogs representing each field. Next came *law and political science*, *geo studies* (in detail, archaeology, geology and geography) and *English*, each field being represented by three blogs in the data. *Culture studies* were represented by one blog about culture studies and one about anthropology. Another two blogs were represented by *biological studies*. Two blogs, the fields of which were ethics and

'complexification studies' were difficult to categorise in any way, so they were grouped together in the category of *other* fields.

The data shows that academic research blogs are being written about a diverse selection of different academic fields. What is more, there seems to be no prevalence of soft science blogs over those of hard science.

4.1.2 Audience and reader characteristics

Analysing audience and reader characteristics adds more value to the analysis of the discourse community that produces and uses academic research blogs. The audience and readers were divided into four categories, *fellow academics*, *lay people*, *like-minded people* and *N/A*, a group for blogs from which neither audience or readers could be determined. The following table illustrates the intended audience of the blogs, that is, who the author seems to have had in mind as a target reader when writing the blog.

Table 7: *Intended audience*

Readers	Frequency	Percentage
Fellow academics	13	43.3
Lay people	10	33.3
Like-minded people	6	20.0
N/A	1	3.3

Most of the academic research blogs (13 out of 30) seem to have been written for fellow academics. This can be determined from the way these blogs use language that only fellow scientists in the field could understand. For example, text like this renders a lay person out of countenance:

- (1) "The changes include fixes to the JavaDoc generation, now outputting proper citations of PhD thesis and books, a fix in the SDFWriter to inherit the IO options from the underlying (used) MDLV2000Writer, restored atom type perception in SMILES parsing if aromaticity is not actively perceived (in line with earlier 1.4.x behavior that unfortunately got broken due to another fix), a fix in the MDLV2000Reader to deal with pseudo atoms with numbers greater than 9 (thanx to John May!), a fix in the sorting of IAtomContainers, and a fix for the elusive bug in the AWTRenderer causing thin bonds (e.g. due to zooming) to become grey" (*Chem-bla-ics: CDK 1.4.5: the changes, the authors, and the reviewers*)

One third of the blogs appeared to be directed to lay people, that is, people that have no prior knowledge of the field but are, nevertheless, interested in the subject matter. In these instances, the blog author may have had a scientific topic, but wrote about it in the way that a person without any prior knowledge could understand. For instance, the author of the blog *Aetiology* discussed the science of biology in this manner. The topics of the entries could be of scientific nature, but the author explained the more scientific parts and terms in more detail through adding explanations in parentheses (see example 19, p. 46) or through links to explanatory Web pages. There were six blogs that were directed at like-minded people. These blogs had a scientific, controversial topic, and the blog author had a very strong opinion of the topic and was more likely trying to confirm the readers to keep on having the same opinion as opposed to trying to get the opposing side to agree with the author.

It has to be taken into account that this division of intended audience is only suggestive. Most of the research blogs seemed to have been written to a very heterogeneous audience from friends and relatives to co-researchers and random people interested in the topic. However, in this analysis, each blog was only coded for one primary audience.

Next, the readers of the blogs were placed in the same categories as the audience above. The reader characteristics were determined from the comments they had written in the comment sections of each blog. Table 9 below shows the division of readers into the four categories.

Table 8: Readers

Readers	Frequency	Percentage
Fellow academics	19	63.3
Lay people	2	6.7
Like-minded people	1	3.3
N/A	8	26.7

The majority of the blogs (19 out of 30) seems to have been read by fellow academics. Only two blogs were, according to who comments on them, read by people who did not seem to have prior knowledge of the relevant academic field. One blog was mainly read and commented by like-minded people, and the rest, 8 blogs had so few or vague comments that it was impossible to determine any general characteristics of the readers.

It seems that there is a considerable discrepancy between the intended audience and the actual readers. Even though most of the academic blogs are primarily directed to groups other than fellow academics, it is the fellow academics that are the most active in openly establishing themselves as the readers of a blog. The result of the reader

analysis cannot, however, be put very much emphasis on, because it does not tell anything about those readers who choose not to comment on the blog entries, and this group of readers can be speculated to be very substantial. The reason why fellow academics are so well represented in the data may be that fellow scientists might have more courage to comment on an academic blog, whereas lay people may be more timid to do so, thinking they do not know enough about the topic to add anything valuable to the conversation.

4.1.3 Interaction between the author and the audience

Before turning to the communicative purpose of the academic blogs, I will briefly discuss the amount and form of interaction between the blog author and its audience. Below is a table that presents a summary of the ways the blogger interacts with the audience.

Table 9: Interaction between blog author and audience

Interaction	Frequency	Percentage
Audience references in blog entries	17	56.7
Author takes part in conversation	17	56.7
No interaction	7	23.3

In seventeen blogs, the author addressed their audience in one way or another. The blogger might, for example, ask for the audience's opinion or ideas on some matter:

(2) "See why I liked this book and told you to go look at it? I always love ideas for class exercises. Got any more? then pass them along!" (*Academic Cog: Not what you thought*)

Another way the author might be interacting with the audience is to take part in the conversation happening in the comments section of each entry. This happened in 17 blogs. In seven blogs, there was no interaction between the author and the audience. The readers may have asked or otherwise commented on the blog entries, but they did not lead to any kind of dialogue with the author.

So, the majority, with 23 blogs out of 30, there was some interaction and dialogue between the blogger and the audience.

4.2 COMMUNICATIVE PURPOSES

Perhaps the most vital part of any genre analysis is to define the communicative purpose of the discourse that is being analysed. According to Swales (1990: 4), “[t]he principal criterial feature that turns a collection of communicative events into a genre is some shared set of communicative purposes”. Therefore, I will now turn to identifying the set of communicative purposes that the academic research blogs in my data share. Upon close examination of the blogs, four general purposes arose, and each blog was assigned to one of the four categories of *invisible college*, *diary about academic life*, *popularising science*, *soap box* and *sharing information with fellow academics*. The primary purposes of the blogs are summarised in table 11 below.

Table 10: Primary purpose

Primary purpose	Frequency	Percentage
Invisible college	12	40.0
Diary about academic life	6	20.0
Popularising science	6	20.0
Soap box	6	20.0

The data shows that the most common communicative purpose for an academic research blog was *invisible college* with 12 blogs belonging to this category. Following closely were *diary about academic life*, *popularising science* and *soap box* (6 blogs each). The following subsections are dedicated to explaining in more detail the characteristics of the set of four communicative purposes represented in the data.

4.2.1 Invisible college

Invisible college blogs were blogs that were created and maintained to keep in touch with fellow academics and to share and ask for opinions and ideas on scientific topics relevant to the blog author’s research. A blog with a purpose of creating an *invisible college*, could, for instance, be a blog where the author picks the latest trends and developments in their research field and brings their own perspective to them, or a group blog sharing information on upcoming conferences and calls for papers. In some of the blogs, the goal of creating an *invisible college* network was explicitly presented:

- (3) The overarching goal of this blog is to build a stronger feminist law prof community across scholarly subject areas. (*Feminist Law Professors*)

In the political science blogs, the topic could also be a current event, which the author then analyses scientifically and prompts the readers to share their views on the subject. For example, in a blog called *Brian's Coffeehouse*, which according to header information is "Commentary on the Politics, History and Culture of the Middle East and Central Asia", the author talks about the Arab upheaval in this manner:

- (4) As a historian, I find it unsurprising that a revolution would traverse multiple phases, as that is simply what often happens. This is especially true when there is no ready made united opposition to assume the helm. (*Brian's Coffeehouse: The Next Phase*)

Some of the blogs in the *invisible college* category were mostly dedicated to talking about the author's own research and topics connected to it. There were entries about the technicalities of writing a paper, or about some striking findings the author has come up with, or recaps of seminars or conferences the author has attended.

The language of the *invisible college* blogs signalled that the blogs were directed to people who have a significant amount of knowledge in the scientific field of the author. This could be detected from the use of abbreviations and equations or word choices inherent to the specific field. The following excerpts show samples of language from the blogs that perhaps would not make a lot of sense to a lay reader:

- (5) And yes, as you say, there are also other sp²-like hexagonal boron nitride. In other areas such as oxides, the interest is not so much a new allotrope but the interaction that arises when you combine different materials in a heterostructure for new effects. (*All that matters: The two sides of promoting materials science*)
- (6) They've observed elliptic flow at center-of-mass-energy/nucleon-pair $s_{NN}=2.76$ TeV. For those who were sceptical, the quark-gluon plasma is still a strongly-coupled liquid, with very low shear viscosity, at LHC energies – more than an order of magnitude larger than the $s_{NN}=200$ GeV at RHIC. (*Musings: Alice!*)

It seems that most of the blogs the purpose of which was to create an *invisible college* managed to get fellow academics to share their opinions, thoughts and ideas with the author and other readers. The fellow researchers took part in the conversation in the comments

section of the entries, and, as the following examples illustrate, the comments are written by people who clearly know a thing or two about the topic:

- (7) Renormalized and IR-fixed pure QED does not need a UV completion. It works fine at $L \rightarrow \infty$ without limitation. (Comment from *Musings: Alice!*)
- (8) I know that English comp teachers (like me) find it impossible to teach even one overload, while folks in other areas were (past tense) able to work double loads (mostly teaching online courses). That's a real disconnect. (Comment from *Confessions of a Community College Dean: Overloads*)

However, not all the blog entries in each blog belonging to the category of *invisible college* were strictly about science or academic life, or written in a scientific language opaque to a lay reader. Many of the blogs in this group included entries that were written with a lay reader in mind, without the goal of stirring up academic interaction between peers. These entries were, nonetheless, outnumbered by those that were written with the purpose of creating scientific conversation between fellow academics.

4.2.2 Diary about academic life

Like *invisible college* blogs, the category of diary about academic life often has the blog author's research and academic life as a topic too, but the tone of these blogs is very different from that of *invisible college* blogs. In the six blogs that were identified as diaries about academic life, the entry topics ranged from academic topics such as presenting own research, asking readers for opinions on research problems or dealing with university students to the more traditional diary text, such as talking about shopping, gardening or cats. The texts in this category are much more personal than in the *invisible college* blogs, with more emphasis on the authors' leisure time and family life.

Whereas *invisible college* blogs were informative and neutral, diary blog authors were not afraid of using colloquial language and telling readers what they really think about their academic work. As can be seen from the examples below, the difference in the language is rather stark when compared to that of *invisible college* blogs.

- (9) The second and more significant negative is that it means that I need to deal with students' personal shit. And sometimes that personal shit fucks me up, too. (*Reassigned Time: Office Hours*)
- (10) I have a journal all picked out that I really really want to get published in even though I find them snooty and I hate them. Oh, and they have rejected my articles before. And were

obnoxious. And yet their articles are delicious tasty mental snacks and I want to be listed alongside those big name snooty-mac-snoots. So one of my little mindless tasks is to find their submission guidelines and suggested article length and make a list of biblio stuff to finish. I had forgotten these idiots use Chicago Style. Oh fuck me. (*Academic Cog: Chicago Freakin' Style!?!?!?!?!?!?!?!?*)

- (11) *The Statue Room* is...well, it's...OK, it's pretty horrifying. And not because it's Gothic. Silverfish may be too generous for this novel. (*The Little Professor: Yhe Statue Room*)

The blogs in this category seem to be used for the purpose of venting off, letting out steam that builds up in academic work. In a blog, the author can express how annoyed they are with the students they teach or with the tediousness of scientific publishing. Bloggers can also talk about their research in honest remarks without having to analytically process and mould their thoughts into academically sound opinions. For example, the author of *The Little Professor* blog hardly will describe the gothic novel *The Statue Room* as "pretty horrifying" in an academic article, but can freely do so in a blog.

The diary blogs have attracted a more diverse reader base than the *invisible college* blogs, where most of the comments were written by fellow researchers in the same field of science. In the diary blogs, the commenting readers seem to be a mix of fellow academics in the field, fellow academics in another field, lay readers friends and relatives.

- (12) Oh no not Chicago Style. I loathe it, loathe it. I had to deal with it for an article that got published a few months ago. (Comment from *Academic Cog: Chicago Freakin' Style!?!?!?!?!?!?!?!?*)

- (13) I've read a few books like this. Not being a trained English professor, I sort of assumed that the problem was me and my inability to properly interpret the text. It is comforting to learn that I was wrong. (Comment from *The Little Professor: The Statue Room (vol.II)*)

- (14) This last "Dear Freshman" is amazingly accurate & well stated. Thanks for making me smile. **A fellow teacher. (Comment from *Learning Curves: Transparency of Grading*)

A noteworthy finding about the *diaries about academic life* is that of the six pseudonymous blogs in the data sample of the study, as many as four belong to this group. Only two of the six blogs belonging to the group of *diaries about academic life* provide the real full name of the author.

4.2.3 Popularising science

Six of the blogs in the sample were analysed to be written with the purpose of popularising science. The blogs in this group were clearly dedicated to presenting and explaining science to any reader interested in the topic, regardless of whether they had a university degree. The topics of the blogs ranged from discussing how microbial world reaches our everyday life to presenting the latest news in biological anthropology and related fields. Most of the blogs in this category included entries about how the blog author's field of science is present in everyday life. For example, a paleontologist dedicated one blog entry to talking about a fossil hominoid depicted on a beer label (example 15), and a biomedical engineer explained what happens in a compulsive shopper's brain (example 16):

(15) I just made what what may be the most amazing discovery of the century at a local booze emporium. Dogfish Head brewing company makes a beer whose label is adorned with Jay Matternes's reconstruction of an upright *Ardipithecus ramidus*. Note that the left foot grasps the earth with it's ape-like big toe. (*Lawn Chair Anthropology: ARDIPITHECUS BEER!!!*)

(16) I have a belly full of turkey and corn-bread dressing, and I'm about to go shopping! Doorbusters, Holiday sales, 50% off, Buy-one-get-one-free, FREE (with a new 2-year contract of course)... If these words get your heart racing and the endorphins flowing, perhaps we should take a look at a rather obscure phenomenon known as Oniomania - better known as compulsive buying. (*From The Lab Bench, S(erotonin) is for shopping*)

In addition, science popularisers tended to talk about their own everyday life relatively often in their blogs, writing about taking their dog to the vet or about noticing how their garden has changed over the years. What has to be noted, though, is that even these entries include some explaining of science: A trip to the vet inspires the biomedical blogger to tell the audience what a hookworm is and what it does, and the wearing and sinking of the archeologist's garden path brings about thoughts on the passing of time.

The language used in the blogs with the purpose of popularising science suggests that the blog authors are writing their blogs for both scientists and non-scientists. Some blog entries were clearly aimed at lay readers, such as the thorough account of witches in example 17 or the word *nano* in example 18 suggest:

(17) A witch is early modern (and late medieval) Europe's projection of its own ultimate imagined Other, the inversion of all the values that society most appreciated - a witch

worshipped the Devil instead of God, engaged in sexual orgies in lieu of continence, killed cattle, made people sick and messed with the weather, rather than living as a useful member of the community. The witch was the infertile, jealous woman who poisoned infants, instead of the nurturing Christian mother. A witch was the parody of all that early modern Europeans believed held their fragile world together. (*Somerville Historian: Witches*)

- (18) 'Nano' describes the science of what happens when we shrink things down to one billionth of a meter, 100 times smaller than the diameter of an average strand of human hair. (*From The Lab Bench: Why Tiny Cars Make Headlines*)

Sometimes the blog entry was written in a way that the blog author gave additional, explanatory information to the lay readers either in parenthesis (example 19) or by hyperlinking the scientific terms to their respective Wikipedia pages (example 20):

- (19) Now, Varicella zoster (chickenpox) isn't on their list as far as "select agents," but secretions from a person thought to have or diagnosed with chickenpox would be considered a category B agent (moderate risk of harm). (*Aetiology: Chickenpox parties – just a Facebook friend away*)
- (20) In animals with a tail, a number of caudal vertebrae continue off the set of fused vertebrae called the sacrum. Humans and other apes don't have true tails but a coccyx, a small clump of tiny, fused vertebral segments. (*Lawn Chair Anthropology: ARDIPITHECUS BEER!!!*)

There were, however, entries in the blogs which were relatively unintelligible for lay readers, even though the blogger might have stated that the blog was meant to present science to non-scientists. For example, a psychology PhD student writing her blog about studying in Edinburgh talks about her dissertation process in a relatively difficult language:

- (21) I'm busy checking Freesurfer data (cortical and subcortical segmentations) and brain segmentations from FSL for another VBM analysis. (*Martina's PhD Blog: The paradox of my brain analyzing my brain*).

When analysing the comments sections of the blogs in this category, it is evident that both scientists and non-scientists are reading the blogs. Fellow scientists tend to offer critique or additional viewpoints on the subject matter (example 22), whereas lay readers ask for more information or thank the author for insightful posts (example 23):

- (22) Then, there is the Xenotropic murine leukemia virus-related virus link with human prostate cancer, though German researchers were unable to detect the virus (US researchers

did) (Comment from *Aetiology: Does bestiality increase your risk of penile cancer?*)

(23) What do the cave sediments usually consist of? Windblown leaves and dust would not penetrate far inside. Rubble falling off the roof? Animal faeces? (Comment from *Aardvarchaelogy: Recent Archeomags*)

So, academic research blogs, the purpose of which is to popularise science, tend to be written for and read by both scientists and non-scientists. The blog authors in this group are clearly trying to make science more accessible to non-scientists, which can be seen in the explanatory writing style of the blogs, and also in that the popularising bloggers readily communicate with their audience by often referring to them in the texts (in four of the six blogs in the category) and answering to the comments given to their texts (in four of the six blogs).

4.2.4 Soap box

Like the categories of *diary about academic life* and *popularising science*, there were six blogs belonging to the group called *soap box*. This is a category which is based on Walker's (2006: 4) notion of blogs that are maintained by what she calls "public intellectuals". According to Walker, (*ibid.*) "many academic bloggers use their blogs as a platform for political debate based on theories of political science, feminism, discourse and media analysis and so on". In this category, the blogs have as their general topic in addition to the subjects Walker lists above subject matters such as pseudoscience, evolution and ethics. The blog authors in this group are not afraid of giving their opinions on even the most debatable and controversial subjects, with the idea of provoking discussion. The topics of the blog entries in a blog usually deal with a hot potato which is somehow connected to the author's research field, which gives the blogging scholar a higher specialist status on the subject matter, making it, when combined with the heavily additudinal way of writing, seemingly difficult for the readers to disagree with the author. The authors often present their thought world and general viewpoints already in the beginning on the about page:

(24) I'm your basic white heterosexual middle-class cisgender non-disabled US-citizen raised-mainline-Protestant tall thin college-educated married male from a happy intact nuclear family background, and I'm sure it shows (*Debitage*)

(25) Evolution, development, and random biological ejaculations from godless liberal (*Pharyngula*)

The language in the soap box blogs does not differ very much from that of popularising science in terms of specificity. Most of the blog entries are intelligible to lay readers, but, at the same time, accurate enough for fellow academics. Where popularising blogs tended to use hyperlinks to other resources as explanatory additions, soap box bloggers use them as proofs and sources for their opinions and ideas on a given topic:

- (26) Scientists have made a new strain of bird flu that most likely could spread between humans, triggering a pandemic if it were released. A misguided project, or a good idea? How should we handle dual use research where merely knowing something can be risky, yet this information can be relevant for reducing other risks? (*Practical Ethics: Ferretting out fearsome flu: should we make pandemic bird flu viruses?*)
- (27) It turns out that Bradshaw appears to be a homeopath, as I discovered when I first heard of this movie several months ago. At least, that's what she lists her occupation as in her political campaign contributions" (*Respectful Insolence: Anti-vaccine propaganda lands in New York City this weekend*)

The starkest difference between *popularising science* blogs and *soap box* science blogs is that the *soap box* blogs are written in a very attitudinal manner, while popularising blogs tended to be of a more neutral nature. Examples 28 and 29 are samples of text where the author is not afraid to show their negative attitude towards the people they are referring to in the blog. In example 28, the blog author diminishes the scientific value of the hypotheses presented in *Medical Hypothesis* (MH) by placing selected words in quotes. The author in example 29 uses more direct ways of giving their opinion on Newt Gingrich, calling him a philanderer, sleazebag, odious creature and even a pretentious moral monster.

- (28) MH has "distinguished" itself from other medical and scientific journals in its willingness to publish things such as a "hypothesis" by autism quacks Mark and David Geier that provided the basis for their use of chemical castration to treat "vaccine-induced" autism, a "hypothesis" claiming that antiperspirants cause breast cancer, a "hypothesis" by prominent anti-vaccine activists that thimerosal in vaccines causes autism, as well as other hypotheses claiming that masturbation is a treatment for nasal congestion or that high-heeled shoes are linked to schizophrenia. (*Respectful Insolence: The return of the Medical Hypothesis anti-vaccine howler*)
- (29) Finally, Herman Cain has suspended his presidential campaign under a cloud of accusations of sexual harassment and adultery, clearing the field for...Newt Gingrich? Philanderer and sleazebag? This is not a step up in ethics. At last, I understand something, though. Gingrich has been an odious creature throughout his entire political career; a venal, pretentious moral monster. (*Pharyngula: Cain is out*)

In many of the blogs in this category, the authors encourage people from all walks of life to take part in the debate in the comments section. As the topics of the blog entries tend to be topics that anyone with an interest in current events has some kind of opinion of, the readers who comment on the posts seem to be both fellow scientists and lay readers, although in most of the cases it was impossible to know if a commenter was a scientist in the field or just an avid newspaper follower, as can be seen in examples 30 and 31 below.

(30) Administration of such ingroup-bias-suppressors may help, for instance, in reducing racial discrepancies in sentencing. Of course, we need to be mindful of the intervention's effect in other domains such as empathy (one of the interesting aspects of the findings Crockett presented was how many aspects of moral thinking were affected by oxytocin and other interventions - I wonder if it is even possible to perform truly domain-specific moral manipulation); but perhaps, if we are uncertain about the importance/desirability of those other traits but certain that bias (in this case, at least) is undesirable, then the intervention would make sense. *(Comment from Practical Ethics: What Moral Virtues Should We Enhance?)*

(31) One thing I've noticed about the Occupy movement is a lack of a unified platform. I've seen all sorts of ideas coming out of Occupy, ranging from abolishing corporate personhood to reinstating the financial regulations of the 1930s Glass-Steagall Act *(Comment from Pharyngula: A completely justified howl of moral outrage)*

What is interesting about the *soap box* blogs is that even though they are purposely written in a provocative way to stir up discussion on controversial topics, most of the commenting readers share the blog author's opinions. Only in a few entries in the sample there was some heated debate between the readers and sometimes between readers and the audience. Therefore, according to this study, it would seem that the *soap box* blogs might be written in order to share one's specialist opinion with like-minded people.

4.3 STRUCTURAL FEATURES

As mentioned earlier, Freedman and Medway (1994: 8–9) consider the content and form of a genre text to rest upon the social, technical and institutional factors present in the communicative situation. In this section, I will summarise the most prominent features of the structural form of the blogs in the data sample in order to map the form the different factors present in the communicative situation have created. Below is a table representing a summary

of the structural features analysed in this study, and following the table is further analysis of the most common structural characteristics present in the blogs:

Table 11: Structural features

Feature	Frequency	Percentage
Date	30	100.0
Comments allowed	28	93.3
About section	26	86.7
Archives	23	76.7
Images	22	73.3
Tags	22	73.3
Blogrolls	18	60.0
Badges	14	46.7
Videos	11	36.7
Advertisements	9	30.0
Calendar	5	16.7

Herring *et al.* (2004: 7) remind that the use or non-use of many of the structural characteristics listed above is determined by the blogging software used to create and maintain the blog. What is common for every blog creation software is that they impose a display of blog entries in reverse chronological order in a one- to three-column format. In addition, different blogging software have many default settings, such as comments and archives, and a novice user might not know how to change these settings. Therefore, the following results of the structural analysis should be taken with a grain of salt.

As can be seen from Table 11 above, all 30 blogs in the sample shared the structural feature of showing the date of publishing for each entry. This feature can be speculated to derive from journals, one of the source genres of blogs. On the other hand, one of the unrecognised rules of blogging is to publish frequently (Herring *et al.* 2004: 2) and adding the date to each entry posted on the blog might be the author's way of helping the readers to stand the pace of Internet publishing.

As for the comments, there are some notable differences between the findings of Herring *et al.* (2004) and those of this study. Herring *et al.* found out that comments were allowed in only 43.0 per cent of their blogs, whereas in the blogs analysed in the present study there were only two blogs where commenting was restricted. In the data sample in the study of Herring *et al.*, the average entry received 0.3 comments, whereas the mean number of comments per post in this study was 8.2. Based on their findings, Herring *et al.* suggest that the degree of interactivity in blogs might not be as high as previous research had claimed. The

present study, however, shows that the academic research blogs in the sample are relatively interactional. The difference might be explained by the fact that the authors of the majority of the academic research blogs in this study encouraged their readers to take part in the discussion, and most of the topics were likely to be more debate provoking than an average general blog. Nonetheless, one explanation for the notable difference is undoubtedly the fact that in 2011, the year the data was collected for the present study, many more people are reading blogs than in 2004 when the data sample for the study of Herring *et al.* was compiled.

Most of the blogs in this study had some internal information organisation system. 23 bloggers out of 30 had organised their blog entries in an archive according to the date each entry was published. Different kinds of tags, categories and other folksonomy-based organising systems were employed in 22 blogs. This labelling of content was often used to separate the academic content from the non-academic. For example, in the blog called *Little Professor*, entries with academic content were tagged with the labels *academic*, *teaching* or *science*, whereas entries about the blog author's everyday life were put in the categories such as *Friday cat blogging*, *scattered musings* or *food and drink*.

The social and interactive nature of online media was apparent in the way academic bloggers wanted to show their affiliations to different people, groups, institutions, societies or causes. There were two prominent ways to do this, and those were by attaching badges and creating blogrolls to one's blog. Badges are small icons in the boilerplate areas of the blog, and they advertise group affiliation. For example, the blog called *Practical Ethics* had badges for *Oxford Uehiro: Centre for Practical Ethics*, *Institute for Science and Ethics*, *The Oxford Centre for Neuroethics*, *Future of Humanity Institute*, and *Oxford Martin School* (see Figure 3 below). Some blogs had badges that showed other than academic affiliation, for instance the blog *Learning Curves* had badges for *I <3 wool* (a badge for a knitting community site), *Keep away from children* (an image from a warning from kitchen matches), and a counter for the author's progress in knitting mittens (see Figure 4). Attaching badges was clearly less popular among scholars who blog than among the general bloggers in the study of Herring *et al.* (2004). In my data sample, 14 blogs, making up 46.7 per cent, had badges, whereas the findings of Herring *et al.* (2004: 7) show that in their data, as much as 69.0 per cent of the blogs included badges in their sidebar, header or footer.



Figure 4: Badges on Practical Ethics



Figure 5: Badges on Learning Curves

Another popular way of expressing one's affiliation on a blog is to create a blogroll to the sidebar of the blog. A blogroll is a list of links to other blogs, and these listed blogs usually are blogs that the authors themselves read. In the sample of this study, 18 blogs out of 30 had such a list of links to other blogs in their sidebars, making blogrolls a little more popular a way of showing affiliation than attaching badges to one's blog.

The data in the study of Herring *et al.* (2004: 7) suggested that blogs are relatively image-poor with 58.6 per cent of the blog home pages containing pictures and only 9.2 per cent of the separate blog entries having images, and therefore not making as much use of the multimedia potential of the Web as other Web genres do. In the present study, the number of blogs that had images in the blog entries was considerably higher, with only eight blogs containing no images in the entries collected from the first page, making the percentual value of blogs with images in their entries as high as 73.3%. In addition, 11 blogs out of 30 had embedded video links to the blog entries, a feature that Herring *et al.* (*ibid.*) had not analysed. The substantial difference in the amount of multimedia usage between the blogs in the study of Herring *et al.* and the present study cannot be considered to suggest that academic blogs would make more use of images and videos than general blogs. Herring *et al.* (2004: 8–9) point out that the low amount of images in their blog sample might be due to the fact that in data collection, they only accepted blogs that had textual content in the most recent entry. What is more, in the six years that separate Herring *et al.*'s study from the present analysis, major technological development has taken place in the blogging software, making it easier to

add an image or embed a video in the entries, and the amount of digital cameras and mobile phones with a camera has increased drastically, leading to bloggers having greater access to multimedia content.

According to Blood (2002), a prototypical blog entry is centered around a link to external content. In addition, Luzón (2008: 75) argues that links are essential in academic blogs because academic blog authors use their blogs to situate themselves in a disciplinary blogging community. Therefore, I will turn my focus on looking at what kind of hyperlinks the blog entries in my data sample contain. While Luzón (2008: 79), in his analysis on linking in academic blogs, categorised the links in academic blogs into 9 different groups ranging from link to social media to link to audio files for the post, I narrowed the categories in my study to four different types of links that could be detected in the data. These types are *reference*, *self-reference*, *explanation* and *entertainment*, and their distribution in the data is summarised below in table 13:

Table 12: Links

Link type	Frequency	Percentage
Reference	591	56.1
Self-reference	147	14.0
Explanation	302	28.7
Entertainment	13	1.2
Total	1053	100

The majority of the links in the sample belonged to the category called *reference*. These were links that were used to show the source on which the blog author builds their arguments and opinions on. These sources were, for example, online newspapers or other news sites, abstracts or full versions of research articles, or video interviews. Example 32 illustrates two reference links, the first one linking to a news source *The Denver Channel*, and the second to an abstract of a research article. In 9 of the 30 blogs, the blog authors had decided to give reference links in the formal academic style at the end of the blog entry, as example 33 shows below.

(32) This study was described in news articles that appeared around the time of its release, but, as always, I'll go to the source, an article in JAMA entitled Neuron Number and Size in Prefrontal Cortex of Children With Autism. (*Respectful Insolence: Picking up one's marbles and going home*)

(33) Reference:
Wilson, C., Johansson, G., Pourkabirian, A., Simoen, M., Johansson, J., Duty, T., Nori, F., & Delsing, P. (2011).

Observation of the dynamical Casimir effect in a superconducting circuit *Nature*, 479 (7373), 376–379 DOI: [10.1038/nature10561](https://doi.org/10.1038/nature10561) (*All that matters: Plucking light out of space*)

Self-reference links, on their part, were links to the blog author's earlier entries in the blog or links to their publications. Of all the links in the data sample, 14.0 per cent represented this category. This is a much higher percentage than its corresponding value presented in the study of Herring *et al.* (2004: 9), which was as low as 5.2%. Below are two examples of *self-reference* links, the first one linking to the blog author's publication and the second to a previous blog entry:

- (34) The great news for me, in this week, has been the acceptance of my paper of QCD at finite temperature in *Physical Review C* (see [here](#)). (*The Gauge Connection: QCD at finite temperature*)
- (35) As with my previous [reflections](#) post, I've posted rather late so if your interested in what happened at the conference check out the following write-ups (*Physiological Computing: Reflections on Quantified Self Europe 2011*)

The third link type, *explanation*, contained links to resources that gave relevant further information on the subject matter, but did not work as a reference for the source for the author's arguments or opinions. This was the second most popular link type in the data, with 28.7 per cent of the links belonging to this category. Example 36 shows how explanation links were used as a way to provide additional information of institutes, and example 20 (see p. 46) illustrates how some words bear links to their Wikipedia sites to offer more information for those who need it.

- (36) Funding for space missions in the UK was split from the [Science and Technology Facilities Council](#) to the the [UK Space Agency](#) earlier this year. Very roughly, UKSA will fund the missions themselves all the way through to the processing of data, while STFC will fund the science that comes from analysing the data. (*Leaves on the Line: STFC and UKSA*)

The fourth category, *entertainment*, contains links to resources with pure entertainment value. The topics of the linked sources often had nothing to do with the general purpose or subject matters of the blogs. Usually these entertainment links were funny videos, images or sites the blog author had found somewhere on the Internet and had decided to share them with the readers, as in the example 37 below:

- (37) The greatest of all Wikipedia lists: *_List of dates predicted for apocalyptic events_*, from some gloomy Assyrian to the Heat Death of the Universe! (*The Cranky Professor: The end of the world is (always) nigh!*)

The average amount of links per blog entry was 4.07, which is strikingly higher than the corresponding number in the study of Herring *et al.* (2004: 9), which was 0.65. So, it would seem that the use of the Web's medium-feature of hyperlinking is more exploited in the academic research blogs than in general blogs. I will consider the possible reasons for this difference in the following Discussion section.

4.4 LINGUISTIC FEATURES

This section is dedicated to carrying out the sixth step of Bhatia's (1993) genre analysis. As was stated in the methods and material chapter, Bhatia (*ibid.*) suggests a genre analyst to perform the linguistic analysis of a genre text on one or more of the linguistic realisation levels of analysis of lexico-grammatical features, analysis of text-patterning or textualisation, and structural interpretation of the text-genre. I will now turn into analysing the lexico-grammatical features of the corpus of academic blog entries with comparison to the respective findings in Biber's (1988) study. In my analysis, I will focus on quantifying some of the linguistic features that are considered to be typical of academic discourse, such as nominalisations and heavy usage of prepositions, in order to find out if there is a difference in the use of lexico-grammatical realisations between academic blogs and academic prose. I will also count the instances of features that are untypical of academic prose, to see if these features manifest themselves in academic research blogs. To check the validity of the results of the analysis of my corpus, I will compare my findings to those of Stuart's (2006) in his similar study.

It can be hypothesised that the most visible linguistic differences between academic blogs and academic prose are in the degree of formality and interactionality, and also in informational density of the text. Therefore, the linguistic features chosen to be analysed show some aspects of the degree of formality, interactionality and informational density. The linguistic features analysed here that are typically considered to represent interactional discourse are *first and second person pronouns*, *hedges*, *amplifiers*, *possibility modals* and *private verbs*. Features that point to more informational discourse are, for example, *nominalisations* and *prepositions*, which are also quantified in this study. To see if the language in the academic blogs is more informal than in academic prose, I analysed the

frequencies of *emphatics*. The list of all the linguistic characteristics coded for in this study is represented in Table 14 below:

Table 13: Comparison of linguistic features in academic blogs with Biber's study of academic prose (1988: 255)

Feature	Total Frequency (Academic blogs)	Mean Frequency (Academic blogs)	Mean Frequency (Academic prose)
Amplifiers	361	2.1	1.4
Analytic negation	703	4.1	4.3
Conjuncts	484	2.8	3.0
Discourse particles	393	2.3	0.0
Downtoners	341	2.0	2.5
Emphatics	1135	6.6	3.6
Existential <i>there</i>	321	1.9	1.8
First person pronouns	3737	21.9	5.7
Hedges	131	0.8	0.2
Indefinite pronouns	2472	14.6	0.2
Necessity modals	332	1.9	2.2
Nominalisation	4009	23.5	35.8
Possibility modals	846	4.9	5.6
Predictive modals	652	3.8	3.7
Prepositions	17451	102.1	139.5
Pronoun <i>it</i>	1342	7.9	5.9
Second person pronouns	671	3.9	0.2
<i>Seem/appear</i>	177	1.0	1.0
Synthetic negation	218	1.3	1.3
Third person pronoun	3619	21.2	11.5
Word length	-	4.7	4.8

In order to make it possible to compare the amount of the linguistic features studied with Biber's (1988) earlier study, mean frequencies for these features were calculated. The calculations were done by normalising the frequencies of the linguistic features to a text length of 1000 words. For example, there were 332 necessity modals in the corpus, and the mean frequency was attained by dividing the number of the necessity modals by the total number of words in the corpus, and then multiplying the result by 1000:

$$332 / 170953 \times 1000 \approx 1.942$$

In addition, the linguistic features studied here and the listings of which words constitute a certain feature (see Appendix) are taken from Biber's study (1988) to allow for comparison between this study and that of Biber's.

As is apparent from the table above, there were many linguistic features that showed no apparent difference in frequency between academic blogs and academic prose. These include, among others, the frequency of use of modals. In addition, analytic and

synthetic negations, hedges, existential *there* and pronoun *it* were roughly as common in academic blogs as in academic prose.

Differences between the two corpora do, however, arise when looking at the mean frequencies of features that signal interactional discourse. Amplifiers, that is, extent or degree adverbs or adverbials that are used to intensify meaning (e.g. *very*, *absolutely*, *completely*) were significantly more common in the corpus of academic blogs than in academic prose, but even more considerable variation was found when quantifying the mean frequencies of pronouns. Both first and second person pronouns, in addition to indefinite pronouns, were exceedingly more frequently used in academic blogs than in academic prose, which can be interpreted as showing a more interactional stance to the audience. What is more, discourse particles were relatively common in academic blog texts, whereas there were no instances of them in academic prose.

An interesting finding in the present study was that also third person pronouns were clearly more frequently used in academic research blogs than in academic prose. In his earlier study on the subject, Stuart (2006) found that only first and second person pronouns showed marked variation between his corpus of academic blogs and that of Biber's (1988). In my study, however, third person pronouns were basically as common in the blogs as first person pronouns, which might suggest a certain shift of discourse is taking place in the blogosphere from writing pure personal diaries to more diversified collection of discourse styles.

According to Biber (1986), nominalisations and prepositions occur in instances where highly abstract information is conveyed, and therefore they are common in academic prose. The results of the calculating of these two linguistic features showed that nominalisations and prepositions are not as common in academic blogs as in academic prose. However, when compared to Biber's (1988) analysis of the amount of nominalisation across different registers, it is apparent that even though the mean frequency of nominalisations is not as high in academic blogs as it is in academic prose, it is still more common to use nominalisations in academic blogs than other written genres, excluding press editorials.

Emphatics (e.g. *really*, *so + adj.*, *such a*) are characteristic of informal and colloquial discourse (Chafe 1985), and it is perhaps no surprise that they are clearly more common in academic blogs than academic prose. Interestingly though, Stuart's (2006) study did not show remarkable differences in the use of emphatics between the two corpora, but in the present study the variation between academic blogs and academic prose is more pronounced.

5 DISCUSSION

The purpose of this study was to get an overall picture of the genre of academic research blogs. Academic research blogs were considered an interesting topic worthy of closer examination because they seem to be created at the intersection of personal and public, and informal and formal. Blogs are descendants of diaries, which are commonly regarded as a very personal kind of text, but blogs, on the other hand, are often shared with anyone with an Internet access. What is more, blogs have often adopted the informal writing style of diaries and other similar texts in the social media, but researchers are accustomed to use a very formal language when writing about their research in research articles or monographs. Therefore, it is worth analysing what kind of discourse is created when scholars start to write about their research in the form of blogs.

To tackle the task of mapping the genre of academic research blogs, I formulated the following research questions for this study:

1. What are the generic features of academic research blogs?
2. How do academic blogs differ from blogs in general?
3. How does the language of academic research blogs differ from that of academic discourse in general?
4. Where do academic research blogs fit in the genre ecology of academic discourse?

The generic features of academic research blogs were examined by analysing the discourse community that produces and uses the blogs, the communicative purpose of the blogs, and their prominent structural and linguistic features. In the following sections, I will discuss the results of my analysis of these four elements of the blogs studied. In addition, I will contemplate the possible differences and similarities detected in the results between academic research blogs and blogs or academic discourse in general. Finally, I will discuss the position of academic research blogs in the genre ecology of academic discourse. Before moving on to discussing the reliability and validity of the present study, I will formulate an overall summary of the answers this study provides to the research questions given above.

5.1 DISCOURSE COMMUNITY

When analysing the discourse communities that produces and uses academic research blogs, characteristics of the blog authors, readers, audience and their interaction was coded for. The

results of the demographical analysis of the blog authors showed that the prototypical blogging scholar is an individual male researcher residing in the U.S., and these findings were very similar to those of Herring *et al.*'s (2004) study on blogs in general. However, features that differentiated academic blogs from blogs in general started to crop up in the further analysis. According to the results, a notable amount of 80 per cent of the bloggers revealed their full real name in the blog, whereas the respective percentage in blogs in general was as low as 31.4 per cent. What is more, blogging scholars were clearly more ready to provide information on their occupation and field of science, including their titles and ranks in the universities. So, the academics who blog seem to be the same type of people as the people who write blogs in general, but the amount of personal information in the academic blog is much higher. This can be speculated to result from various factors. First, as academic blogs tend to be about the work and research of the scholar who writes them, it is often inevitable that occupation-related information is revealed. This is not always the case in general blogs, because they do not automatically deal with the blog author's occupation. Second, one of the underlying purposes of many academic blogs was to disseminate and make known the blog author's research to the wider audience, and this is difficult to do in a credible way if the author does not provide background information on who they are and what and where they are doing their research. The third possible reason for blog authors giving personal and institute-related information about themselves might have to do with the institutional practices the authors carry from the world of academia to the world of blogging. According to Walker (2006: 2–3), the hierarchical structure of academia is very important and always present in the hallways and offices of universities, and evidently its importance is no less in the blogosphere. It seems that the title and rank of the blogging researcher is regarded as one of the most vital pieces of information provided in the blog, perhaps because announcing them explicitly gives more credibility to the blog, and because it simply is a habit that the researchers are accustomed to in their work in the physical world.

One interesting finding in the results in the analysis of the fields of science represented in the blogs was that there were surprisingly many blogs from the hard sciences. According to Stuart (2006: 392–393), hard science blogs are outnumbered by those of the soft sciences because research in hard sciences revolves around scientific inquiry and publishing results, not around presenting one's subjective ideas. However, the present study showed no such prevalence of soft sciences blogs, but different fields were quite equally represented in the data. This result may be evidence for the fact that blog as a medium has evolved during the six years that separate the present study from the Stuart's (2006) study. While blogs in 2006 were not thought to be as suitable for publishing results or scientific inquiry as they were

for writing essays on a person's own ideas, it seems that nowadays it is not unusual for mathematicians to explain in a blog the thought process that led to a certain equation, or a group of physicists to tell about the problems they were having with a particle accelerator.

As for the intended users of the blogs, authors had created and maintained their blogs with different audiences in mind. Almost half of the blogs were aimed at fellow academics, and one third to lay people. Selecting the intended audience surely is one of the most significant decisions a blogging researcher has to do, because it affects the nature and tone of writing, as well as the specificity of the language use in the blog. What was interesting when analysing the comments section in order to establish the reader characteristics, was that there was a stark discrepancy in the amount of blogs written for lay people and blogs commented by lay people. It was evident in the data that most commenting readers were fellow academics, even in the blogs that were mainly written to lay audience. These results, however, do not tell if most of the readers actually are fellow academics, only that most of the *commenting* readers are. Lay people may feel diffident about commenting on an academic research blog, even though they would be avid followers of them. Fellow academics, on their part, may have more courage to share their views and opinions in the comments sections, because they have an expert status on the subject matter because of their education and research. Even though there were not so many blogs that were actively commented on by lay people, it has to be noted that the blogs at least are within the reach of lay people. The affordances of the blog medium brings science closer to the everyman, which in itself is something that the increasingly expensive and password protected scientific journals cannot do. Another thing that sets academic research blogs apart from scientific articles is the affordance of social media to create immediate interaction. In most of the blogs in the sample, there was substantial interaction between the blog author and the readers. The interaction provided by blogs can be used as a bridge to connect science and research with lay people – who, in many instances, are the sponsors of the research through taxation – and also as a connecting bridge between fellow scholars around the world to brainstorm together.

5.2 COMMUNICATIVE PURPOSE

As Swales (1990: 45–46) argues, one of the most important characteristics of a genre is a shared set of communicative events. The results of the present study revealed that the communicative events brought about by the academic blogs in the data set represented the communicative purposes of *invisible college*, *diary about academic life*, *popularising science*

and *soap box*. These four categories can be considered to form a uniform set of communicative purposes, as the underlying reason for communication in all of the four groups is to mediate academic content to wider audience through networked computers in an unofficial manner. Next, I will go through each of the four categories of communicative purposes and discuss the findings of my analysis in more depth.

Invisible college blogs were the most common type of academic blogs in the data sample of this study. The blogs were written and maintained with the purpose of creating an *invisible college*, that is, creating a forum where fellow academics in the field could bond, brainstorm and share research results, views and opinions with each other through networked computers without the limitations and boundaries of the physical world. Undoubtedly scholars and their research can suffer from the physical isolation from other researchers in the field, so perhaps the reason why *invisible college* blogs were the most common type is that researchers want to find new ways to connect with fellow scholars. Intellectual property rights and other ownership issues might prevent scholars from officially asking for co-operation from fellow academics, but stirring up unofficial and general discussion on some scientific issue through the means of blogging may be a tempting alternative. Furthermore, the popularity of *invisible college* blogs might account for the surprisingly many academic blogs from the hard sciences. Hard science research is often conducted by groups of researchers, and hard science scholars may therefore be accustomed to share their research-related problems and ideas with other people in the group. Therefore it might not be such a big step for hard science scholars to start publishing their thoughts on science in a blog.

But do academic blogs succeed in creating an *invisible college*? In most of the blogs in this category, the discussion between the author and the readers stayed on a very general level, and there are at least two obstacles for creating a useful brainstorming environment in blogs. First, blogging scholars tended to present their research on a very general and vague level, which undoubtedly is resulted from IPR laws and ownership issues. The full research and its results is published in a scientific article or as a book, and until that, the particulars of the research are kept to oneself. The second obstacle is the availability of the blogs. Even though the use of the Web and social media is an everyday activity for a young person in the Western Countries, it has to be remembered that the majority of the world population have no access to the Internet (Internet World Stats). Of course, most academics have access to it and are accustomed to using the Web, but blogs are still a relatively new Internet phenomenon, and assumingly not all senior academics have made visiting blogs a part of their academic work. Concluding with these barriers in mind, *invisible college* blogs do show a promising way of connecting with fellow academics around the world. Even though

the communication between scholars stays on a general level, the blogs still offer an arena for scholars to create a network of researchers who are interested in similar research issues, and, it can, in the best case scenario, be a starting point for coming up with a new research topic or idea, or creating a new research group in the future.

There were six blogs in the data set that were analysed as serving the purpose of being *diaries about academic life*. These blogs were not centred around the research of the blog author, but more on providing a diary-like account of the life of a researcher. In some of the blogs in this category the main purpose seemed to be to vent off all the frustration the blog author has with academic work, be it counselling students or trying to get one's article published. It is perhaps no surprise that these blogs were often pseudonymous, which was in all the other categories very rare. Even though pseudonymity can lessen the credibility of a blog, these pseudonymous diaries about academic life can be considered to play a vital role in the set of academic blogs. The diary blogs were used as a means to introduce one's life and work as a researcher, but also as a place to ask for peer support and help in challenging conundrums faced in academic work. For example, one pseudonymous blogger asked for advice on how to deal with a student who is being accused of having sexually assaulted someone and had come to ask for help from his professor (*i.e.* the blogger). These kinds of dilemmas are something that you do not probably want to bring up at coffee break at the department, but can, with the safety of anonymity, be discussed in a blog. So, *diaries about academic life* provide different things to different sets of readers. Lay people get a snapshot of what the life of a researcher is like, and fellow academics can discuss touchy work-related subjects. What is more, the academic diary blogs provide a valve for blog authors through which to release the frustration that can add up at work, and they often get peer support in the form of discussion in the comments section from fellow academics dealing with similar problems.

The third category of communicative purposes present in the data sample was popularising science, and it was represented by six blogs. The main emphasis of these blogs was to make science and research known to lay people. Hence, the topics of the blog entries tended to be rather general, but still have something to do with the blog author's field of science. Furthermore, the language and style of the popularising blogs was very explanatory, with most of the academic and scientific terms explicated to the lay readership. Blog authors in this category, however, had other means too to make science closer to non-scientists. One prominent habit in the blogs was to knit science to everyday life by talking about an everyday phenomenon and explaining the science behind it, or starting off with an account of an event in the blog author's personal life, and afterwards turning the event to an example of science

affecting our everyday life. The comments posted by the readers of the blogs showed that there were both scientists and non-scientists reading the blogs in this category. Even though the principal intended audience was non-scientists, the blog authors did take into consideration the fellow scientists reading the blogs. This was apparent in the way the bloggers wrote their blog entries. The entries were often written in such scientific detail that fellow academics knew, for instance, exactly which virus or bacteria the blogger was talking about, and the explanatory details for the lay readers were often given in parentheses after the exact scientific term or by hyperlinking the term to online encyclopaedia. In short, the results of the present study seem to allow for the interpretation that while *invisible college* blogs aim at stirring up discussion between fellow academics and diary blogs about academic life function as a valve for venting of and discussing the more personal side of academic work, the blogs in this third category are used as vessels for making science more attainable for lay people and integrating it to the everyday life.

The blogs in the fourth category, *soap box*, differ from the blogs in the other categories not so much in their topics but in the style they discuss these topics. As the name of the category suggests, the blog authors concentrated on giving their opinions on various debatable subjects in their academic field. The blog texts in this group were mostly very opinionated and far from neutral. However, it was not very clear in the data whether the *soap box* blogs were written in order to stir up debate on the subject matter and make people on the opposing side to re-evaluate their stance, or whether the soap box blogs were, like *diaries about academic life*, merely vehicles for the blog authors to vent off and have their say on a subject that is close to their heart and interests. Most of the comments made by readers of these blogs were from like-minded people, that is, people that share the blog author's stances, views and opinions on the subject. Even though the subject matters and the dealing of them in the blogs could have prompted a very heated debate, there very rarely were instances of any kind of dispute.

The present study has shown that academic blogs can be divided into four categories which reflect their communicative purpose. As mentioned above, the underlying communicative purpose of academic blogs was, according to the data sample of this study, to mediate unofficial academic content through the Web to a wide set of people. The blogs were, however, divided into the four subcategories according to the differences in the form and style of the blogs. These differences were brought about by the differing intended audiences of the blogs, and the differing aspects of academia the blog authors wanted to convey to their selected intended audience. *Invisible college* blogs were mostly about the research the blog authors were conducting, and the issues concerned with the research in

hand were shared with fellow academics. Blogs that were interpreted as *diaries about academic life* were not so much centred around the research itself, but more on the life and work in academic environments, and the intended audience was more varied. *Popularising science* blogs, on the other hand, were more concerned with the blog author's research issues, but they were opened up to the non-academic audience instead of fellow academics in the field. The blogs in the last category, *soap box*, were perhaps the most difficult to analyse and categorise, because the intended audience was sometimes very hard to determine, and because the blog authors did not necessarily write that much about their own research or even the field of it. However, what made these blogs academic ones, and not, for example political blogs, was that the blog authors did ground their views and opinions on their first-hand knowledge of science and scientific research. Therefore, these blogs were considered to work for the underlying purpose of mediating scientific content. While all four categories represented different kinds of academic blogs, it could be seen that the *soap box* blogs and *diary blogs about academic life* were both written in a subjective and opinionated style, but the difference between the two groups was that the *soap box* bloggers tended to keep their blogs relatively impersonal, whereas *diaries about academic life* handled very personal issues. *Invisible college* blogs and *popularising science* blogs were similar in that the tone of the blogs in these categories was relatively neutral and objective but differences arised in the intended audiences of these blogs. *Invisible college* blogs were primarily written for fellow academics, while the target audience of the *popularising science* blogs was lay readers with an interest in but no academic knowledge of science.

5.3 STRUCTURAL FEATURES

In the analysis of the structural features of the academic blogs in the data sample, selected structural features were identified and quantified. The results revealed that all of the blogs in the sample showed the date of publishing for each entry, and the majority of blogs had allowed commenting, had a section where the blog author was introduced, organised the blog entries in archives with tags, and had images in the entries. So far, these surface features did not prove any significant difference between the structural features of academic blogs and general blogs. Variation was found, when compared to the study of Herring *et al.* (2004) in the commenting and multimedia habits, with blogs in the present study being more highly interactive and containing more images. These results, however, cannot be interpreted as highlighting a difference between general blogs and academic blogs, because the variation can also result from the changes in Internet behaviour and development in IT technology during

the six years that separate this study from that of Herring *et al.*'s (*ibid.*). In the latter part of 2000's, more and more people have found their way into the blogosphere, which might account for the higher interactivity level of the blogs in my data, and the rising number of digital cameras and speedier data transfer between gadgets may be the reason for the larger amount of image material in my data.

When analysing the ways the academic blogs expressed affiliation, it was found that adding badges to one's blog was less common in academic blogs than in the general blogs analysed in the study of Herring *et al.* (2004). However, the results showed that the academic bloggers were, nonetheless, keen on showing their affiliations to certain communities. Instead of badges, this was mainly done through hyperlinking. The majority of blogs in the sample had a sidebar blogroll, a list of blogs the blog author reads or recommends to readers. The blogs in the blogrolls were often similar to the linking blog, that is, they too were academic blogs, and thus the blogrolls could be, in many cases, considered a way of forming a precursor of an *invisible college*.

What is more, the entries themselves contained a larger amount of hyperlinks than the general blogs in the study of Herring *et al.* (2004). These in-post links were found to link to resources that provided background information and facts on which the blog author based their discussions, be it a link to a research unit Web page, a news site or an academic research article. One of the reasons for the difference between the popularity of linking between general and academic blogs could be that academics are used to explicitly inform the readers the sources and works cited in scientific research, and this habit is strongly present in the academic blogs, too. In fact, in many blogs the authors had put all the sources cited in the blog entry at the end of the post in formal academic citation format.

The affordance of linking can be considered the most important feature that separates academic blogging from other academic discourse. As Luzón (2009: 85) suggests, the multifunctionality and versatility of blogs as a scholarly communication medium is mostly brought about by linking, and it is this multifunctionality and versatility that the printed or spoken academic discourse lack. Through linking, blog authors can integrate and situate themselves in an online community and academic discourse community, learn new ways to construct knowledge, and publicise their own work by adding links to one's own research articles or project Web sites. It is difficult to come up with any affordance provided by other means of academic communication that could serve as many purposes as hyperlinking.

To conclude with, academic research blogs seemed to share the same structural feature as general blogs, but the main characteristic that separates academic blogs from other academic communication and blogs in general is the heavy use of linking. The form of an

academic research blog can be considered to rest upon the following factors: As a technical factor, the default settings of the blogging software often account for the ordering of blog texts into reverse chronological order in a one- to three-column format. Another technical factor present in the communicative situation is the affordance of blogging softwares to allow adding links, images, videos, comments and other multimedia features, which allow for the interactivity and multifunctionality of the blogs. In addition to the technical factors, institutional factors present in the communicative situation shape the form of the genre text. Citation practices inherent in academic discourse are migrated into academic research blogs too, but they often have a slightly different form in the blogs. Instead of a formal bibliography, academic bloggers prefer adding in-post links to resources that serve the same purposes as the formal citations in academic research articles and books.

5.4 LINGUISTIC FEATURES

A corpus-based study was conducted to identify and quantify the linguistic features characteristic of academic research blogs. The linguistic analysis focused on quantifying the degree of formality, interactionality and informational density. When compared to the respective findings of Biber's (1988) study on academic discourse, the results showed that there were many lexico-grammatical features that were equally frequent in academic blogs and academic discourse as a whole. However, the degree of interactivity and informality was found to be much higher in academic blogs than in academic discourse in general. The higher degree of interactivity may be the result of the technical affordances of social media to offer the readers a chance to respond to communication in real time. The blog texts are written in an interactive style because the blog authors can expect their readers to answer to their questions and comments and give feedback because it is made simple and easy for them. In comparison, a scientific article cannot be as interactive as a blog because there is no special medium through which readers of the articles could communicate with the authors as quickly and as easily as through blogs. Surely a reader of an article can contact the author via post or email, but the communication is not instant as considerable time has lapsed between the finishing of the article and its publication. The interactivity and informality of the blog texts also point to one of the antecedents of blogs. Blogs are thought to have their origins in handwritten personal diaries, and the somewhat colloquial and informal language of academic blogs can be considered to be inherited from this offline antecedent of the blogs.

5.5 ACADEMIC RESEARCH BLOG AS AN ACADEMIC GENRE

The present study has provided a preliminary view of the genre of academic research blogs, which makes it possible to hypothesise where academic blogs are situated among other genres of academic discourse. As mentioned in section 2.2.2, the world of academic genres can be organised in various ways, for instance in hierarchies, genre chains, genre sets and networks of genres. This section deals with discussing the place of academic research blogs in the genre ecology of academic discourse in order to understand the relationship and links between the genre of academic blogs and other genres of academic discourse.

When organising academic genres into hierarchies according to how much each genre is respected and valued, it is apparent that regardless of the field of research, academic research blogs are not situated at the top of the genre hierarchy. Swales (2004: 13–18) goes through differences in the order of the most valued academic genres in different disciplines (see also section 2.2.2), but similar to all of the genres that are placed in the key slots of hierarchies is that they all convey new research in a formal and thought-of style and in a respected forum (e.g. research articles in top-ranked journals, published monographs, peer-reviewed conference presentation papers). Academic blogs, on their part, have typically not been peer-reviewed in any way or written in a formal style, and most of the blogs only present research that has already been published elsewhere so it is not entirely new knowledge. What is more, a blog somewhere on the Web can hardly be considered a respected forum for presenting research. To sum up, the place of academic research blogs in the hierarchy of academic genres cannot be considered to be very high in any discipline.

Another perspective of the organisation of academic genres is to put them in a chronological order. These successions of antecedent genres are called genre chains (see section 2.2.2), and, like genre hierarchies, are often different in different disciplines. Some genres are necessary antecedents of following genres, but many genres are optional links in genre chains. Academic research blogs can be considered one of these optional links, as hardly any academic genre-text has to be followed or must be preceded by a blog text. For example, a research article is often preceded by a grant proposal, but instances where a blog text is a pre-requisite for a research article are few if not non-existent. However, academic research blogs can serve an important function as a genre that is used continually as a place for free form doodling and brainstorming of ideas and to refine the thought process of the researcher. Therefore, it can be argued that academic research blogs cannot be situated in a specific chronological point in a genre chain, as their production is often continual and ongoing through all the links in the genre chain.

The formation of academic genres in genre sets means grouping genres in networks of genres that an individual encounters in their institutional and occupational practice (see section 2.2.2). As genre sets have to do with the academic life and work of an individual, there is no clear-cut template of a basic set of academic genres, but often they contain research articles, dissertations, conference presentations and posters and so on. With the increasing popularity of academic blogging, the academic research blog genre is included in the genre set of the ever-growing number of scientists who maintain their own academic blog or follow the blogs of fellow scientists. In the process of acculturation of a researcher, the genre set of the individual tends to expand from course work of a graduate student through preliminary research papers and research articles to the dissertation of the doctoral student. Academic research blogs are an interesting genre in that they can be a steady part of the researcher's genre set from the beginning of the career to the end whereas the understanding of some genres in the set, e.g. the dissertation, can change a lot during the academic career.

The last perspective of the organisation of academic genres discussed here is genre networks. Genre networks take into account the intertextuality of different genres, that is, the way new genres evolve from earlier genres, thus sharing some generic features with each other and forming a network of inter-connected genres (see section 2.2.2). This approach to genre constellations seems to be suitable in situating the genre of academic research blogs in the genre ecology of academic discourse, because academic blogs are a relatively new genre and they have not yet evolved too far from their antecedent genres. For instance, the results of the present study showed that the authors of academic research blogs often continue to use the formal citation system of research prose. Based on the results of this study, it can be argued that while academic research blogs have adopted several generic features from more informal genres, they still share many generic characteristics with other academic genres and can therefore be considered a part of the genre network of academic discourse.

5.7 RELIABILITY AND VALIDITY

Being a relatively small-scale study, the results and implications of the present research should be taken with the proviso that the conclusions made in the study are only suggestive. One of the reasons for this is that the sample size is relatively small, with only 259 blog texts from 30 blogs, forming a corpus of 135 000 words. The corpus can be considered small when compared to that of Stuart's (2006) study, who had compiled a corpus of 16 million words. Therefore, the results attained from the material cannot be extrapolated to reflect all the

academic blogs in the blogosphere. Furthermore, the experience, knowledge, time and other resources of the researcher herself were somewhat restricted.

In analysing the discourse community that uses and produces academic research blogs, difficulties were faced in multiple points. First, the sample of 30 blog authors is not very representative, and therefore the analysis only yielded much generalised results. Furthermore, the whole reader base was impossible to determine, so the users analysed were only those readers who ventured to comment on the blogs. The analysis of the interaction between the blog author and readers could have been taken much further, for example by looking at the interaction from a different angle and applying, for instance, the theory of face threatening acts by Brown and Levinson (1987).

In addition, the analysis of the communicative purposes of the blogs could have been developed further for example by looking at the mechanisms used in expressing the newness of information and conveying shared knowledge in a text type that is directed both to fellow academics and lay readers, or by analysing the way the blog authors publicised their own research from the viewpoint of promotional language. It can be argued that no blog is created and maintained simply in order to serve one communicative purpose, and it was also apparent in the data sample of this study. Therefore, it has to be pointed out that the classification of the set of communicative purposes present in the academic blogs is a crude generalisation of an abundance of varying minor purposes.

The analysis of the structural features of academic blogs was straightforward *per se*, but the comparison to the structural features quantified in the study of Herring *et al.* (2004) was problematic because of the time lapse between the two studies. The blogosphere, and computer-mediated communication as a whole, is a fast evolving and changing area of communication, as improvements and development in information technology instantly affect the form of discourse on the Internet. Therefore, it is difficult to determine whether the higher amount of hyperlinking and multimedia usage is a generic feature of academic blogs or whether it is just the form blogs in general have adopted during the recent years.

Texts, particularly electronic texts that allow for hyperlinking, are very multidimensional constructs, and to fully analyse and understand them, one would have to look at them from multiple perspectives. The linguistic analysis in this study remained very narrow for a number of reasons, the main reason being the problematic size of the text corpus. 135 000 words formed too small a corpus to allow for a reliable and proper statistical analysis, but too big a corpus for an individual researcher to tag all the words according to their parts of speech. Therefore, the lexical analysis was only carried out in the lexico-grammatical level, even though the analysis of clausal features and move structures would

surely have derived very interesting results and offered a more multifaceted understanding of the genre text.

However, the results and implications of the present study do serve the purpose of this research. In many sections, the analysis could have gone much further, but the goal of this study was to provide an overall picture of the genre of academic research blogs. Based on the analysis of the study, a general notion of how academic research blogs present themselves at the time of the research can be formed. What is more, the blogs chosen for this study were selected randomly from a relatively comprehensive set of academic blogs, which increases the possibility of the data sample being representative enough to allow for making apt conclusions about the genre of academic research blogs.

6 CONCLUSION

The purpose of this study was to get an overall picture of the genre of academic research blogs. Academic research blogs were considered an interesting topic worthy of closer examination because they seem to be created at the intersection of personal and public, and informal and formal. Blogs are descendants of diaries, which are commonly regarded a very personal kind of text, but blogs, on the other hand, are often shared with anyone with an Internet access. What is more, blogs have often adopted the informal writing style of diaries and texts in the social media, but researchers are accustomed to use a very formal language when writing about their research in research articles or monographs. The objective of this study was to find out what kind of text is produced through the interplay of personal, public, formal and informal antecedent texts. The present study has been a preliminary attempt at getting a snapshot of the genre of academic research blogs at the time of the research.

In order to study the genre of academic research blogs, 30 academic blogs written in English were collected, from which a corpus of 259 blog entries and 135 000 words was compiled. Genre analysis was conducted following the guidelines of Swales (1990) and Bhatia (1993). The discourse communities that use and produce the blogs were identified, the communicative purposes of academic research blogs were analysed, and their structural and linguistic features were identified and quantified. This analysis provided a preliminary idea of the form of the genre of academic research blogs, which was the main objective of the study.

The main findings of the present study are summarised below in accordance with the research questions that have formed the backbone of this research. The first goal of this study was to determine the generic features of academic blogs. In order to do this, an analysis of the discourse community, communicative purpose and structural and linguistic features was carried out. It was found out that a prototypical academic research blog in the data sample was created and maintained by an individual male researcher in the U.S., and the blog was mostly written for and read by fellow academics in the field. The prototype of academic research blog had as its communicative purpose to form an *invisible college*, an informal forum for fellow academics to share ideas and opinions between each other. The structural analysis of the blogs showed that academic research blogs had a considerably conventional blog structure, but the degree of hyperlinking was notably high. The study of the linguistic features prominent in the data revealed that the language of academic research blogs was highly interactional and rather informal, having still some characteristics of academic language.

The second research question dealt with the variation between the generic features of academic research blogs and blogs in general. The comparison of the blogs in the data sample and the results of the study of Herring *et al.* (2004) showed that the major differences between academic research blogs and general ones was that unlike the often pseudonymous general bloggers, the authors of academic blogs in most cases revealed their real identity in their blogs. In addition, the amount of multimedia features and hyperlinking was much higher in academic blogs than general blogs.

The third research question was posed in order to find out how the language and communication in academic research blogs differs from that of academic discourse in general, and what role do academic research blogs play in the colony of academic genres. The results of the present study suggest that while the discourse in academic blogs can be identified as a form of academic discourse, it is still much more interactional and informal than the discourse in academic settings in general.

Finally, academic blogs can be considered to form a distinct part in the constellation of academic genres. The ranking in the genre hierarchies of academic discourse was considered low because of the lack of peer-reviewing and other ways of ensuring the academic quality of the blog texts. It was thought impossible to pinpoint academic research blogs in a chronological point in the genre chains of academic discourse, as the producing of blog texts often is an ongoing process through the succession of other genres. In addition, academic blogs were found out to be an optional link in the genre chain and also in the genre sets of individual researchers. Lastly, academic research blogs were considered to have their place in the genre network of academic discourse because of the clear inter-connectedness between the generic features of academic prose and academic blogs evident in the results of the analysis conducted in the study.

Further analysis could be conducted on a variety of issues. For example, an analysis of the users and not only producers of academic blogs could be carried out in order to for a more wide-ranging view of the discourse community. When it comes to analysing the communicative purpose of the blogs, it could be developed further for example by looking at the mechanisms used in expressing the newness of information and conveying shared knowledge in a text type that is directed both to fellow academics and lay readers, or by analysing the way the blog authors publicised their own research from the viewpoint of promotional language. Texts, particularly electronic texts that allow for hyperlinking, are very multidimensional constructs, and to fully analyse and understand them, one would have to look at them from multiple perspectives. Therefore, expanding the linguistic analysis from

lexico-grammatical level to the analysis of clausal features and move structures would surely yield very interesting results and offer a more multifaceted understanding of the genre text.

REFERENCES

Primary sources:

Aardvarchaeology. [online] Available at <<http://scienceblogs.com/aardvarchaeology/>> [Accessed 14 December 2011].

Academic cog. [online] Available at <<http://academiccog.blogspot.com/>> [Accessed 28 November 2011].

Aetiology. [online] Available at <<http://scienceblogs.com/aetiology/>> [Accessed 5 December 2011].

Alice Richardson. [online] Available at <<http://ise.canberra.edu.au/alicer/>> [Accessed 8 December 2011].

All that matters. [online] Available at <<http://blog.joerg.heber.name/>> [Accessed 30 November 2011].

Brian's coffeehouse. [online] Available at <<http://bjulrich.blogspot.com/>> [Accessed 2 December 2011].

Chem-bla-ics. [online] Available at <<http://chem-bla-ics.blogspot.com/>> [Accessed 30 November 2011].

Confessions of a community college dean. [online] Available at <<http://suburbdad.blogspot.com/>> [Accessed 8 December 2011].

Cranky professor. [online] Available at <<http://crankyprofessor.com/>> [Accessed 1 December 2011].

Debitage. [online] Available at <<http://blog.debitage.net/>> [Accessed 2 December 2011].

Empirical legal studies. [online] Available at <<http://www.elsblog.org/>> [Accessed 14 December 2011].

Feminist law professors. [online] Available at <<http://feministlawprofessors.com/>> [Accessed 29 November 2011].

From the lab bench. [online] Available at <http://blogs.nature.com/from_the_lab_bench/> [Accessed 15 December 2011].

Gauge connection. [online] Available at <<http://marcofrasca.wordpress.com/>> [Accessed 15 December 2011].

In medias res. [online] Available at <<http://inmedias.blogspot.com/>> [Accessed 29 November 2011].

Lawn chair anthropology. [online] Available at <<http://lawnchairanthropology.blogspot.com/>> [Accessed 15 December 2011].

Learning curves. [online] Available at <<http://learningcurves.blogspot.com/>> [Accessed 8 December 2011].

Leaves on the line. [online] Available at <<http://www.andrewjaffe.net/blog/>> [Accessed 7 December 2011].

Little professor. [online] Available at <<http://littleprofessor.typepad.com/>> [Accessed 1 December 2011].

Martina's PhD. [online] Available at <<http://martinasphdblog.blogspot.com/>> [Accessed 13 December 2011].

Musings. [online] Available at <<http://golem.ph.utexas.edu/~distler/blog/>> [Accessed 7 December 2011].

Paul White's pvc. [online] Available at <<http://uspace.shef.ac.uk/blogs/PVCLandT>> [Accessed 12 December 2011].

Pharyngula. [online] Available at <<http://pharyngula.org/>> [Accessed 5 December 2011].

Physiological computing. [online] Available at <<http://www.physiologicalcomputing.net/>> [Accessed 12 December 2011].

Plektix. [online] Available at <<http://plektix.fieldofscience.com/>> [Accessed 14 December 2011].

Practical ethics. [online] Available at <<http://blog.practicaethics.ox.ac.uk/>> [Accessed 13 December 2011].

Profgrrrrrl. [online] Available at <<http://profgrrrrrl.com/>> [Accessed 25 November 2011].

Reassigned time. [online] Available at <<http://reassignedtime.wordpress.com/>> [Accessed 28 November 2011].

Respectful insolence. [online] Available at <<http://scienceblogs.com/insolence/>> [Accessed 24 November 2011].

Somerville historian. [online] Available at <<http://www.somervillehistorian.blogspot.com/>> [Accessed 13 December 2011].

Secondary sources:

Androutsopoulos, J. (2006) Introduction: Sociolinguistics and computer-mediated communication. *Journal of Sociolinguistics*, 10/4, pp. 419–438.

Askehave, I. & Nielsen, A. E. (2005) Digital genres: a challenge to traditional genre theory. *Information Technology & People*, 18/2, pp. 120–141.

- Bates, M.J. and Lu, S. (1997) An exploratory profile of personal home pages: content, design, metaphors. *Online and CD-ROM Review*, 21/6, pp. 331–340.
- Bazerman, C. (1988) *Shaping Written Knowledge: The Genre and Activity of the Experimental Article in Science*. Madison: The University of Wisconsin Press.
- Bazerman, C. (1994) Systems of genres and the enhancement of social intentions. In Freedman, A. & Medway, P. (eds.) *Genre and New Rhetoric*. London: Taylor and Francis, pp. 79–101.
- Becher, T. & Trowler, P. (2001) *Academic tribes and territories: Intellectual inquiry and the cultures of disciplines*. Milton Keynes: SRHE and Open University Press.
- Bhatia, V. K. (1993) *Analysing Genre: Language Use in Professional Settings*. London: Longman.
- Bhatia, V. K. (2002a) Applied genre analysis: a multi-perspective model. *Ibérica*, 4, pp. 3–19.
- Bhatia, V. K. (2002b) A generic view of academic discourse. In Flowerdew, J. (Ed.) *Academic discourse* (pp. 21–39). Harlow: Longman.
- Biber, D. (1986) Spoken and written textual dimensions in English: Resolving the contradictory findings. *Language*, 62, pp. 384–414.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, D. & Conrad, S. (2009) *Register, genre and style*. Cambridge: Cambridge University Press.
- Biber, D., Conrad, S. & Leech, G. (2002) *The Longman student grammar of spoken and written English*. London: Longman.
- Biber, D., Conrad, S., Reppen, R., Byrd, P., Helt, M., Clark, V., Cortes, V., Csomay, E. & Urzua, A. (2004) Representing language use in the University: Analysis of the TOEFL 2000 Spoken & Written Academic Language Corpus (T2K-SWAL). *TOEFL Monograph Series*. Princeton, New Jersey: Educational Testing Service.
- Biber, D. (2006) *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.
- Blood, R. (2000) *Weblogs: A history and perspective*. *Rebecca's Pocket*. [online] Available at <http://www.rebeccablood.net/essays/weblog_history.html> [Accessed 5 December 2011].
- Blood, R. (2002) *The Weblog Handbook: Practical Advice on Creating and Maintaining Your Blog*. Cambridge, MA: Perseus.
- Brown, P. & Levinson, S. (1987) *Politeness*. Cambridge: Cambridge University Press.
- Chafe, W. (1985) Linguistic differences produced by differences between speaking and writing. In Olson, D., Torrance, N. & Hildyard, A. (eds.) *Literature, Language and Learning: The Nature and Consequences of Reading and Writing* (pp. 105–123). Cambridge: Cambridge University Press.

- Chandler, D. (1998) *Personal homepages and the construction of identities on the web*. [online] Available at <www.aber.ac.uk/media/Documents/short/webident.html> [Accessed 9 December 2011].
- Charles, M. (2003) 'This mystery. . .': A corpus-based study of the use of nouns to construct stance in theses from two contrasting disciplines. *Journal of English for Academic Purposes* 2/4, pp. 313–326.
- Charles, M. (2006a) Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines. *English for Specific Purposes* 25/3, pp. 310–331.
- Charles, M. (2006b) The Construction of Stance in Reporting Clauses: A Cross-disciplinary Study of Theses. *Applied Linguistics* 27/3, pp. 492–518.
- Charles, M. (2007) Argument or evidence? Disciplinary variation in the use of the Noun that pattern in stance construction. *English for Specific Purposes* 26/2, pp. 203–218.
- Conrad, S. (1996) Investigating academic texts with corpus-based techniques: an example from biology. *Linguistics and Education*, 8, pp. 299–326.
- Conrad, S. (2001) Variation among disciplinary texts: a comparison of textbooks and journal articles in biology and history. In Conrad, S. & Biber, D. (eds.) *Variation in English: multi-dimensional studies* (pp. 94–107) London: Longman.
- Coxhead, A. (2000) A new academic word list. *TESOL Quarterly*, 34, pp. 213–238.
- Crowston, K., & Williams, M. (2000). Reproduced and emergent genres of communication on the World-Wide Web. *The Information Society*, 16/3, pp. 201–216.
- Crystal, D. (2001) *Language and the Internet*. Cambridge: Cambridge University Press.
- Davies, J. & Merchant, G. (2007) Looking from the inside out: Academic blogging as new literacy. In Knobel, M. & Lankshear, C. (eds.) *A new literacies sampler* (pp. 167–198). New York: Peter Lang.
- Devitt, A. (1991) Intertextuality in tax accounting: Generic, referential, and functional. In Bazerman, C & Paradis, J. (Eds.), *Textual dynamics of the professions: Historical and contemporary studies of writing in professional communities* (pp. 336–380). Madison: University of Wisconsin Press.
- Dillon, A. & Gushrowski, B.A. (2000) Genre and the web: is the personal home page the first uniquely digital genre? *Journal of The American Society for Information Science*, 51/2, pp. 202–205.
- Efimova, L., & de Moor, A. (2005) Beyond personal webpublishing: An exploratory study of conversational blogging practices. *Proceedings of the 38th Hawai'i International Conference on System Sciences (HICSS'05)*. Los Alamitos: IEEE Press.
- Efimova, L., & Hendrick, S. (2005). *In search for a virtual settlement: An exploration of weblog community boundaries*. [online] Paper presented at the Second International

Conference on Communities and Technologies 05, Milan, Italy. Available at <<https://doc.telin.nl/dsweb/Get/Document-46041/>> [Accessed 12 December 2011].

- Emigh, W., & Herring, S. C. (2005) Collaborative authoring on the Web: A genre analysis of online encyclopedias. *Proceedings of the 38th Hawai'i International Conference on System Sciences (HICSS'05)*. Los Alamitos: IEEE Press.
- Erickson, T. (2000) Making sense of computer-mediated communication (CMC): conversations as genres, CMC systems as genre ecologies. *Proceedings of the 33rd Hawai'i International Conference on System Sciences (HICSS'00)*. Los Alamitos: IEEE Press.
- Freedman, A. Medway, P. (eds.) (1994) *Genre and the New Rhetoric*. London: Taylor and Francis.
- Flowerdew, J. (Ed.) (2002) *Academic discourse*. London: Longman.
- Groom, N. (2005) Pattern and meaning across genres and disciplines: An exploratory study. *Journal of English for Academic Purposes* 4/3, pp. 257–277.
- Halavais, A. (2006) Scholarly blogging: Moving towards the visible college. In Bruns, A & Jacobs, J. (eds.) *Uses of Blogs* (pp. 117–126). New York: Peter Lang.
- Halliday, M. A. K. & Martin, J. R. (1993) *Writing Science: Literacy and Discursive Power*. London: The Farmer Press.
- Herring, S.C. (ed.) (1996) *Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives*. Amsterdam, John Benjamins.
- Herring, S. C. (2001) Computer-mediated discourse. In Schiffrin, D, Tannen, D. & Hamilton, H. (eds.), *The Handbook of Discourse Analysis* (pp. 612–634). Oxford: Blackwell Publishers.
- Herring, S. C., Kouper, I., Paolillo, J. C., Scheidt, L. A., Tyworth, M., Welsch, P. et al. (2005) Conversations in the blogosphere: An analysis "from the bottom-up". *Proceedings of the 38th Hawaii International Conference on System Sciences (HICSS'05)*. Los Alamitos: IEEE Press.
- Herring, S. C., Scheidt, L. A., Bonus, S., & Wright, E. (2004) Bridging the gap: A genre analysis of weblogs. *Proceedings of the 37th Hawaii International Conference on System Sciences (HICSS'04)*. Los Alamitos: IEEE Press.
- Hiltunen, T. (2010) *Grammar and Disciplinary Culture: A Corpus Based Study*. PhD thesis. Department of English, University of Helsinki.
- Hsu, C. L., & Lin, J. C. C. (2008) Acceptance of blog usage: The roles of technology acceptance, social influence and knowledge sharing motivation. *Information and Management*, 45/1, pp. 65–74.
- Hyland, K. (2001) Humble servants of the discipline? Self-mention in research articles. *English for Specific Purposes*, 20/3, pp- 207–226.

- Hyland, K. (2002a) Directives: argument and engagement in academic writing. *Applied Linguistics*, 23/2, pp. 215–239.
- Hyland, K. (2002b) Options of identity in academic writing. *ELT Journal*, 56/4, pp. 351–358.
- Hyland, K. (2009) *Academic discourse*. London: Continuum.
- Hyland, K. & Tse, P. (2009) Academic lexis and disciplinary practice: corpus evidence for specificity. *International Journal of English Studies*, 9/2, pp. 111–130.
- Internet World Stats. *Internet World Stats: Usage and Population Statistics*. [online] Miniwatts Marketing Group. Available at < <http://www.internetworldstats.com/stats.htm>>. [Accessed 12 March 2012].
- Kjellberg, S. (2010) I am a blogging researcher: Motivations for blogging in a scholarly context. [online] *First Monday*, 15/8. Available at <<http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2962/2580>>. [Accessed 19 March 2012].
- Luzón, M. J. (2009). Scholarly hyperwriting: The function of links in academic weblogs. *Journal of the American Society for Information Science and Technology*, 60/1, pp. 75–89.
- Miller, C.R. (1984) Genre as social action. *Quarterly Journal of Speech*, 70, pp. 151–167.
- Mortensen, T. & Walker, J. (2002) Blogging thoughts: Personal publication as an online research tool. In Morrison, A. (ed.) *Researching ICTs in Context* (pp. 247–279). Oslo, Norway: InterMedia Report.
- Räsänen, C. (1999) *The conference forum as a system of genres*. Gothenberg: Acta Universitatis Gothoburgensis.
- Shepherd, M. & Watters, C.R. (1998) The evolution of cybergenres. *Proceedings of the 31st Annual Hawaii International Conference on System Sciences (HICSS '98)* (pp. 97–109). Los Alamitos: IEEE Press pp. 97–109.
- Stuart, K. (2006) Towards an analysis of academic weblogs. *Revista Alicantina de Estudios Ingleses*, 19, pp. 387–404.
- Swales, J. (1990) *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- Thurlow, C., Lengel, L. & Tomic, A. (2004) *Computer Mediated Communication: Social Interaction and the Internet*. London: SAGE Publications.
- Todorov, T. (1990) *Genres in discourse*. Translated from French by Catherine Porter. Cambridge: Cambridge University Press.
- Walker, J. (2006) Blogging from inside the ivory tower. In Bruns, A. & Jacobs, J. (eds.) *Uses of Blogs* (pp. 127–138). New York: Peter Lang.

Wikipedia: Blog. *Wikipedia, The Free Encyclopedia*. [online] Wikimedia Foundation, Inc.
Available at <<http://en.wikipedia.org/wiki/Blog>>. [Accessed 16 January 2012].

Yates, J. & Orlikowski, W. J. (1992) Genres of organizational communication: A structurational approach to studying communications and media. *Academy of Management Review*, 17, pp. 299–326.

APPENDIX

LEXICAL ITEMS FOR LINGUISTIC FEATURES ANALYSED IN THE STUDY

Below is a list of the linguistic features analysed in the study. The list also indicates the lexical items used in the automatic identification of the linguistic features, presenting the exact forms that were counted as instances of each feature. The lexical items are taken from Biber (1988: 221–245).

Amplifiers:

very | extremely | absolutely | completely | altogether | enormously | entirely | fully | greatly | highly | intensely | perfectly | strongly | thoroughly | totally | utterly

Analytic negation:

not

Conjuncts:

alternatively | altogether | consequently | conversely | eg | e.g. | else | furthermore | hence | however | i.e. | instead | likewise | moreover | namely | nevertheless | nonetheless | notwithstanding | otherwise | rather | similarly | therefore | thus

Discourse particles:

well | now | anyway | anyhow | anyways

Downtoners:

almost | barely | hardly | merely | mildly | nearly | only | partially | partly | practically | scarcely | less | slightly | somewhat

Emphatics:

for sure | a lot | such a | just | really | most | more | real+adj. | so+adj. | do+verb

Existential 'there':

there

First person pronouns:

I | we | me | us | my | our | mine | ours | myself | ourselves

Hedges:

at about | something like | more or less | almost | maybe | sort of | kind of

Indefinite pronouns:

everybody | everyone | somebody | someone | something | everything | anybody | anyone | anything | nobody | no one | no-one | nothing | one | some | both | each | either | neither | all | many | enough | any | much | several | none | little | few

Necessity modals:

must | should | have to | has to | had to

Nominalisation:

Words ending with -tion, -ment, -ness, -ity (plus plural forms)
tion | ment | ness | ity | tions | ments | nesses | ities | sion | sions

Possibility modals:

can | could | may | might

Predictive modals:

will | would | shall

Prepositions:

against | amidst | amid | among | amongst | at | besides | between | by | despite | during | except | for | from | in | into | minus | notwithstanding | of | off | on | onto | opposite | out | per | plus | pro | re | than | through | throughout | thru | to | toward | towards | upon | versus | via | with | within | without

Pronoun IT:

it

Second person pronouns:

you | your | yours | yourself | yourselves

Seem/appear:

seem | appear

Synthetic negation:

no | neither | nor

Third person pronoun:

he | she | it | they | him | her | it | them | his | its | their | hers | theirs | himself | herself | itself | themselves

Word length:

Number of characters divided by number of words